

Titre: Système d'apprentissage multitâche dédié à la segmentation des lésions sombres et claires de la rétine dans les images de fond d'oeil
Title:

Auteur: Clément Payout
Author:

Date: 2018

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Payout, C. (2018). Système d'apprentissage multitâche dédié à la segmentation des lésions sombres et claires de la rétine dans les images de fond d'oeil
Citation: [Master's thesis, École Polytechnique de Montréal]. PolyPublie.
<https://publications.polymtl.ca/3257/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie:
PolyPublie URL: <https://publications.polymtl.ca/3257/>

Directeurs de recherche: Farida Cheriet
Advisors:

Programme: Génie informatique
Program:

UNIVERSITÉ DE MONTRÉAL

SYSTÈME D'APPRENTISSAGE MULTITÂCHE DÉDIÉ À LA SEGMENTATION DES
LÉSIONS SOMBRES ET CLAIRES DE LA RÉTINE DANS LES IMAGES DE FOND
D'OEIL

CLÉMENT PLAYOUT
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLOME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
AOÛT 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

SYSTÈME D'APPRENTISSAGE MULTITÂCHE DÉDIÉ À LA SEGMENTATION DES
LÉSIONS SOMBRES ET CLAIRES DE LA RÉTINE DANS LES IMAGES DE FOND
D'OEIL

présenté par : PLAYOUT Clément

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. HURTUT Thomas, Ph. D., président

Mme CHERIET Farida, Ph. D., membre et directrice de recherche

M. SEOUD Lama, Ph. D., membre

DÉDICACE

À ma famille, mes amis et mes collègues...

REMERCIEMENTS

À l'entame de la rédaction de ce mémoire, je ne soupçonnais pas la nature de la dernière difficulté qui m'attendrait à son terme ; difficulté qu'ont rencontré avant moi plusieurs générations d'étudiants. Elles et moi avons en commun d'avoir été supervisés par le professeur Farida Cheriet et par conséquent d'avoir été confrontés au grand défi de résumer en quelques mots les innombrables bienfaits de cet encadrement et la gratitude que nous lui portons. Ces quelques mots, donc, ne suffiront pas à énumérer l'ensemble des conseils avisés distribués par Farida à toute heure du jour et de la nuit, ses commentaires généreux et éclairés, nos discussions à fort potentiel de motivation... Sans parler de sa gentillesse au quotidien, son support moral, ses encouragements et les petits gâteaux apportés au laboratoire, infaillible source de réconfort ! Ce travail n'aurait pu être mené à terme sans elle, mais c'est aussi une grande leçon d'humanité et de positivisme au quotidien que de travailler dans son équipe.

Je tiens également à remercier Philippe Debanné, associé de recherche au LIV4D, pour ses conseils, sa disponibilité et ses relectures aguerries et pointues. Son aide tout au long de cette maîtrise a été indispensable et m'a permis de viser plus haut. De manière générale, ma gratitude se porte à toute l'équipe du LIV4D, anciens, nouveaux et intermittents, pour nos échanges variés, nos débats parfois sérieux et parfois en roue libre et tous ces partages d'expériences. J'y ai rencontré une passionnante équipe de passionnés, découvert des parcours atypiques et fait la connaissance d'histoires et des cultures provenant du monde entier.

En particulier, j'aimerais remercier Gabriel, collègue et désormais ami de maîtrise, sans qui les difficultés, aussi bien techniques, ~~administratives~~, que scientifiques, auraient été bien plus ardues à franchir. Nos interminables discussions du midi sont et resteront une source d'inspiration de premier choix.

Car je vis avec eux et qu'ils ont dû supporter mes silences butés et mon comportement d'ermite irascible pendant les phases les plus difficiles de la rédaction, je voudrais remercier mes colocataires anciens et nouveaux pour leur présence, patience et leurs bons petits plats : Léa, Maxime, Simon, Estelle, Suzanne, Nathalie, merci de votre bonne humeur à toute épreuve !

Enfin et surtout, je tiens à remercier ma famille pour son support indéfectible, son réconfort moral, ses encouragements et pour m'avoir toujours poussé à cultiver plusieurs passions, à garder un esprit critique, à m'ouvrir au monde et somme toute à devenir un être plus libre.

RÉSUMÉ

Le travail de recherche mené dans le cadre de cette maîtrise porte sur l'exploitation de l'imagerie de la rétine à des fins de diagnostic automatique. Il se concentre sur l'image de fond d'œil, qui donne accès à une représentation en deux dimensions et en couleur de la surface de la rétine. Ces images peuvent présenter des symptômes de maladie, sous forme de lésions ou de déformations des structures anatomiques de la rétine. L'objet de cette maîtrise est de proposer une méthodologie de segmentation simultanée de ces lésions dans l'image de fond d'œil, regroupées en deux catégories : claires ou sombres. Réaliser cette double segmentation de façon simultanée est inédit : la vaste majorité des travaux précédents se concentrant sur un seul type de lésions. Or, du fait des contraintes de temps et de la difficulté que cela représente dans un environnement clinique, il est impossible pour un clinicien de tester la multitude d'algorithmes existants. D'autant plus que lorsqu'un patient se présente pour un examen, le clinicien n'a aucune connaissance a priori sur le type de pathologie et par conséquent sur le type d'algorithme à utiliser. Pour envisager une utilisation clinique, il est donc important de réfléchir à une solution polyvalente, rapide et aisément déployable.

Parallèlement, l'apprentissage profond a démontré sa capacité à s'adapter à de nombreux problèmes de visions par ordinateur et à généraliser ses performances sur des données variées malgré des ensembles d'entraînement parfois restreints. Pour cela, de nouvelles stratégies sont régulièrement proposées, ambitionnant d'extraire toujours mieux les informations issues de la base d'entraînement.

En conséquence, nous nous sommes fixés pour objectif de développer une architecture de réseaux de neurones capable de rechercher toutes les lésions dans une image de fond d'œil. Pour répondre à cet objectif, notre méthodologie s'appuie sur une nouvelle architecture de réseaux de neurones convolutifs reposant sur une structure multitâche entraînée selon une approche hybride faisant appel à de l'apprentissage supervisé et faiblement supervisé. L'architecture se compose d'un encodeur partagé par deux décodeurs spécialisés chacun dans un type de lésions. Ainsi, les mêmes caractéristiques sont extraites par l'encodeur pour les deux décodeurs. Dans un premier temps, le réseau est entraîné avec des régions d'images et la vérité terrain correspondante indiquant les lésions (apprentissage supervisé). Dans un second temps, seul l'encodeur est ré-entraîné avec des images complètes avec une vérité terrain composé d'un simple scalaire indiquant si l'image présente des pathologies ou non, sans préciser leur position et leur type (apprentissage faiblement supervisé).

Cette deuxième étape permet au réseau de voir un nombre bien plus grand d'images, ce genre d'annotations étant bien plus facilement disponible. Elle repose sur l'hypothèse qu'il

est possible d’exploiter une annotation globale pour affiner une prédiction locale, les deux étant corrélées (la présence de lésions étant directement liée à l’état sain ou pathologique d’une image).

Dans ces deux étapes, les images de fond d’œil sont prétraitées de sorte à rehausser les contrastes et corriger les inhomogénéités d’illumination. Une augmentation de données artificielle est opérée pour agrandir la base d’entraînement. L’apprentissage se fait selon une nouvelle fonction de coût visant à maximiser le coefficient Kappa de Cohen.

Les résultats obtenus démontrent l’intérêt de l’approche multitâche, plus performante qu’une architecture monotâche équivalente. D’autre part, l’étape d’entraînement faiblement supervisé permet d’obtenir une amélioration significative des performances, démontrant la capacité du réseau à affiner les caractéristiques extraites en étant entraîné sur une tâche de classification à l’échelle globale, différente mais corrélée à la tâche de segmentation. Finalement, cette architecture obtient des performances comparables avec l’état de l’art, tout en proposant une segmentation simultanée de toutes les lésions.

Dans le cadre de cette maîtrise, nous nous sommes également attachés à étendre ces résultats au-delà du cadre strict de la recherche afin de proposer un déploiement auprès de cliniciens. Ainsi, nous avons introduit le concept d’une plateforme intégrant les résultats d’algorithmes opérant sur le fond d’œil mise à disposition des cliniciens pour leur faciliter le travail d’annotations. Ce projet s’intègre dans le cadre d’une collaboration entre plusieurs équipes d’ingénieurs et de cliniciens réparties à travers le Canada visant à développer une base de données de référence de toutes les lésions rétinienne, afin de pouvoir ultérieurement développer des méthodes robustes de dépistage automatique de la rétinopathie diabétique. À terme, l’architecture d’apprentissage profond multitâche proposée sera intégrée dans une plateforme qui sera distribuée dans différentes institutions hospitalières pour le dépistage et le diagnostic de pathologies oculaires.

ABSTRACT

This work focuses on automatic diagnosis on fundus images, which are a bidimensional representation of the inner structure of the eye. The aim of this master’s thesis is to discuss a solution for an automatic segmentation of the lesions that can be observed in the retina. The proposed methodology regroups those lesions in two categories: red and bright. Obtaining a simultaneous double segmentation is a novel approach; most of the previous works focus on the detection of a single type of lesions. However, due to time constraints and the tedious nature of this work, clinicians usually can not test all the existing methods. Moreover, from a screening perspective, the clinician has no clue a priori on the nature of the pathology he deals with and thus on which algorithm to start with. Therefore, the proposed algorithm requires to be versatile, fast and easily deployable.

Conforted by the recent progresses obtained with machine learning methods (and especially deep learning), we decide to develop a novel convolutional neural network able to segment both types of lesions on fundus images. To reach this goal, our methodology relies on a new multitask architecture, trained on a hybrid method combining weak and normal supervised training.

The architecture relies on hard parameter sharing: two decoders (one per type of lesion) share a single encoder. Therefore, the encoder is trained on deriving an abstract representation of the input image. Those extracted features permit a discrimination between both bright and red lesions. In other words, the encoder is trained on detecting pathological tissues from normal ones. The training is done in two steps. During the first one, the whole architecture is trained with patches, with a groundtruth at a pixel level, which is the typical way of training a segmentation network. The second step consists in weak supervision. Only the encoder is trained with full images and its task is to predict the status of the given image (pathological or healthy), without specifying anything concerning the potential lesions in it (neither location nor type). In this case, the groundtruth is a simple boolean number. This second step allows the network to see a larger number of images: indeed, this type of groundtruth is considerably easier to acquire and already available in large public databases. This step relies on the hypothesis that it is possible to use an annotation at an image level (globally) to enhance the performance at a pixel level (locally). This is an intuitive idea, as the pathological status is directly correlated with the presence of lesions.

In both steps, the network is trained with preprocessed images. The preprocessing consists in a correction of the uneven illumination and a strategy for contrast enhancement that preserves color consistency. The training step aims to optimize a novel cost function based on

the Cohen’s Kappa coefficient.

The results obtained prove that multitasking performs better than a comparable single-task architecture. Moreover, the weakly supervised step is shown to improve significantly the segmentation results. Overall, the architecture obtained performances similar to the state of the art while being more versatile and faster.

In addition to this architecture, we propose the concept of a platform that will allow us to deploy our algorithm on the clinical field. This platform’s aim is to facilitate the labelling process done by clinicians, by providing automatic pre-annotations obtained with our algorithm. This project is part of a collaboration between engineers and clinicians from all over Canada. Its goal is to build a large reference database composed of fully annotated fundus images. Ultimately, this will allow developing robust tools for screening of diabetic retinopathy.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xi
LISTE DES FIGURES	xii
LISTE DES SIGLES ET ABRÉVIATIONS	xiv
CHAPITRE 1 INTRODUCTION	1
1.1 Anatomie de l'œil	2
1.1.1 Pathologies rétinienues	5
1.2 Imagerie diagnostique de la rétine	9
1.3 Objectifs généraux	12
1.4 Plan du mémoire	12
CHAPITRE 2 REVUE DE LITTÉRATURE	14
2.1 Evaluation de la qualité des acquisitions	15
2.2 Méthodes de segmentation	16
2.2.1 Segmentation des marqueurs normaux	17
2.3 Approche globale de la classification et de la gradation de la maladie	22
2.4 Réseaux de neurones et apprentissage machine	24
2.4.1 Le neurone formel	28
2.4.2 Les réseaux de neurones multicouches	30
2.4.3 Métriques d'évaluation de performance	32
2.5 Bases de données publiques	38
CHAPITRE 3 PROBLÉMATIQUE ET OBJECTIFS SPÉCIFIQUES	40

CHAPITRE 4	ARTICLE 1: A NOVEL WEAKLY SUPERVISED MULTITASK ARCHITECTURE FOR RETINAL LESIONS SEGMENTATION ON FUNDUS IMAGES	42
4.1	Introduction	42
4.2	Proposed method	45
4.2.1	Preprocessing	45
4.2.2	Data augmentation	47
4.2.3	Patches extraction	47
4.2.4	Multitask architecture	47
4.3	Experiments	51
4.3.1	Material	51
4.3.2	Training	52
4.3.3	Results	53
4.4	Discussion	57
4.5	Conclusion	59
4.6	Acknowledgement	59
CHAPITRE 5	DÉVELOPPEMENT D'UNE PLATEFORME D'ANNOTATIONS	61
5.1	Spécificités de l'annotation d'images de fond d'œil	61
5.2	Conception d'outils dédiés aux cliniciens	63
5.3	Collecte des données	63
5.4	Choix de programmation	66
CHAPITRE 6	DISCUSSION GÉNÉRALE	68
6.1	Synthèse des travaux	68
6.2	Limitations de la solution proposée	69
6.3	Améliorations futures	70
CHAPITRE 7	CONCLUSION	72
RÉFÉRENCES	73

LISTE DES TABLEAUX

Tableau 2.1	Comparatif des bases de données publiques de fond d’œil	39
Table 4.1	Comparative performance of the proposed approach for bright lesions segmentation	54
Table 4.2	Comparative performance of the proposed approach for red lesions segmentation	55
Table 4.3	Comparative performance on detection at an image-level.	57
Tableau 5.1	Temps d’annotations (moyenne/écart-type) par marqueurs.	62

LISTE DES FIGURES

Figure 1.1	Schéma anatomique de l’œil humain. Image libre de droits , Wikipédia (2018).	3
Figure 1.2	Exemple d’image de fond d’œil, issue de la base de données ROC (Niemmeijer et al. (2010))	5
Figure 1.3	Coupe OCT de la rétine. On y observe la forme sphérique de la surface de la rétine et l’emplacement de la fovéa (lieu d’épaisseur minimale de la rétine) et de la papille (d’où arrive le nerf optique).	11
Figure 2.1	Chaîne de traitement automatique d’une image de fond d’œil.	14
Figure 2.2	Segmentation du disque optique et de l’excavation papillaire.	18
Figure 2.3	Effet <i>Copper Wire</i> sur des artères d’un patient atteint de rétinopathie hypertensive Image libre de droit , Wikipedia (2018)	21
Figure 2.4	Démonstration de la détection de ligne selon la méthode de Nguyen et al. (2013)	22
Figure 2.5	Descente de gradients : convergence vers un minimum global et vers un minimum local	26
Figure 2.6	Trajet de la descente de gradient par la méthode Adam.	28
Figure 2.7	Comparaison entre un réseau de neurone réel et son modèle artificiel .	29
Figure 2.8	Évolution du κ dans un contexte de fort déséquilibre de classes en fonction du rapport VP/FP	38
Figure 4.1	Examples of non diabetic patients exhibiting retinal lesions. In figure 4.1a, the patient exhibits a secondary retinal hemorrhage consecutive to a pancytopenia caused by a myelodysplastic syndrom. In figure 4.1b, the patient recovers from BRVO (Branch Retinal Vein Occlusion) and presents a small microaneurysm near the fovea.	45
Figure 4.2	Simplified representation of the proposed architecture. (a) An encoder block is shared among two decoders (b), each of which is specialized in its class of lesions (red or bright). An additional block (c) is trained on detecting whether the presented patch is pathological or not. During the first step of the training, all branches are trained. During the second step, only (a) and (c) are trained on full images.	48

Figure 4.3	Multi-task architecture. The encoder is shared for the two decoders, trained to segment bright and red lesions, and the generalization block, trained to predict whether the input contains a lesion. By training the generalization block, we also train the encoder.	49
Figure 4.4	At the lowest point of the encoder, large kernel convolutions are simulated by the successions of a 7×1 convolution and a 1×7	50
Figure 4.5	Color preserving preprocessing: the top row shows the original images, bottom row shows the results of the preprocessing step.	53
Figure 4.6	Validation at a pixel level of the model on the two databases DIARETDB1 and e-ophta. We show the result before and after the generalization step done on EyePACS. In all configurations, this step improves the segmentation performance.	56
Figure 4.7	Validation at an image level of the model on the two databases DIARETDB1 and e-ophta. A prediction containing at least one positive pixel is considered as referable.	56
Figure 4.8	Lesion segmentation examples. Figures (a), (c) and (e) are the original images extracted respectively from Messidor, DIARETDB1 and EyePACS. Figure (b) only shows the red lesions detection : it illustrates the failure to segment the large haemorrhage at the bottom as a single region. Some small vessels are also counted as false positives in the vicinity of the OD. In figure (d), two biomarkers detected by our model as red lesions are highlighted. The bottom one is indeed a μA but the top one is a false positive photocoagulation laser mark. despite this specific case, only a few of the many laser marks in this image are classified as lesions, proving that this is not a systematic misclassification. Figure (f) illustrates a successful case of exudate segmentation in a poor quality, very low-contrasted image. Many other approaches would require a prior segmentation of the optic disk while it hardly visible in this case.	60
Figure 5.1	Proposition d'interface utilisateur graphique.	64
Figure 5.2	Organisation des annotations en catégories hiérarchisées.	64
Figure 5.3	Interface d'évaluation de la qualité des images.	65
Figure 5.4	Organigramme de fonctionnement de la base de données proposées. .	65
Figure 5.5	Captures d'écran de l'interface dans son état actuel.	67

LISTE DES SIGLES ET ABRÉVIATIONS

AVR	Rapport Artère/Veine
BoVW	Bags Of Visual Words
CLAHE	Contrast Limited Adaptive Histogram Equalization
CDR	Cup-to-Disc Ratio
CNN	Convolutional Neural Network
DMLA	Dégénérescence Maculaire Liée à l'Âge
DO	Disque Optique
DR	Diabetic Retinopathy
ETDRS	Early Treatment Diabetic Retinopathy Study
HSV	Human Visual System
OMD	Cedème Maculaire Diabétique
OCT	Tomographie en Cohérence Optique
RD	Rétinopathie Diabétique
ROC	Receiver Operating Characteristic
SLO	Scanning Laser Ophthalmoscopy
SURF	Speeded Up Robust Features
SVM	Support Vector Machine

CHAPITRE 1 INTRODUCTION

Pour comprendre les pathologies affectant l'œil et les lésions visibles dans la rétine, il est tout d'abord nécessaire d'introduire les bases anatomiques qui décrivent la structure de l'œil. La vision humaine est le résultat de la focalisation des faisceaux lumineux dans le foyer de l'œil : la rétine. Les rayons lumineux pénétrant parallèlement dans l'œil suivent leur chemin à travers différentes interfaces qui vont venir les focaliser vers le point central de ce tissu photosensible. Réciproquement, en inversant le chemin des rayons, il est donc possible d'observer la rétine. Cette propriété mal-aimée des photographes est d'ailleurs la cause de l'effet "yeux rouges" qui touchent les modèles aux yeux clairs photographiés avec flash. Mais aussi indésirable qu'elle soit sur un plan artistique, la possibilité d'imager un tissu actif de façon non-invasive est une grande opportunité diagnostique. En effet, en plus de la capacité à transcrire la lumière focalisée en signaux électriques, la rétine, tissu métaboliquement très actif, est le siège de nombreux échanges sanguins. Son observation est donc riche en enseignements. La topologie et la morphologie du réseau vasculaire rétinien sont des informations d'autant plus précieuses qu'elles sont facilement accessibles, aussi bien pour le diagnostic de maladies de l'œil, du cerveau que liées à la circulation sanguine. Par ailleurs, de nombreux autres biomarqueurs (lésions, structures anatomiques présentant une anomalie...) sont également observables, indiquant de façon parfois très précoce le développement d'une maladie. Fort de cet intérêt pour l'observation de la rétine, les techniques d'imagerie n'ont cessé de s'affiner, gagnant en résolution et en facilité d'acquisition. En conséquence, la quantité de données à analyser s'est accrue. Le passage à l'acquisition numérique a permis l'essor du développement d'algorithmes de plus en plus sophistiqués d'assistance au diagnostic. De l'évaluation de la qualité de l'acquisition au dépistage automatisé de maladie, en passant par la détection et la segmentation de lésions, l'assistance par ordinateur n'a eu cesse de se multiplier. L'émergence récente des méthodes d'apprentissage profond n'a fait qu'accentuer le phénomène, permettant d'élargir le champs d'applications tout en améliorant, parfois significativement, les performances des algorithmes.

Parallèlement, la prévalence de certaines maladies oculaires (notamment la rétinopathie diabétique) nécessite la mise en place de programme de dépistage à des échelles nationales. Néanmoins, le déploiement d'approches automatisées ne peut se faire sans garantie sur la capacité du système à réagir face à des problèmes très variés. Pour cela, il est donc nécessaire d'avoir une quantité importante de données annotées par des cliniciens, aussi bien à des fins d'apprentissages pour les algorithmes que de vérification de la performance de ceux-ci. Malheureusement, l'annotation de données est une tâche fastidieuse, particulièrement longue et

les bases de données publiques déjà existantes ne correspondent pas aux besoins. En effet, celles-ci se concentrent ou bien sur une annotation globale (un diagnostic par image) ou bien sur une lésion spécifique. Aucune ne propose une annotation de l'ensemble des lésions potentiellement visibles sur la rétine. Cet état de fait a entraîné une multiplication des algorithmes (par lésions et par bases de données). Or, en pratique, un clinicien ne peut pas se permettre de tester autant d'algorithmes et préférera une plateforme intégrée recherchant toutes les lésions dans une image. Ce constat a guidé le déroulé de cette maîtrise.

Dans les sections qui suivent et en guise d'introduction, nous présenterons les notions essentielles d'anatomie de l'œil permettant de contextualiser les travaux de recherche réalisés. Les principales pathologies affectant la rétine seront également décrites, l'accent étant mis sur la rétinopathie diabétique, dont l'impact sociétal est majeur. Pour compléter ce survol clinique des enjeux afférant à la rétine, un bref aperçu des modalités d'imagerie de celle-ci est dressé. En conclusion de ce propos liminaire, nous dresserons les objectifs généraux ayant guidé l'orientation de nos travaux.

1.1 Anatomie de l'œil

L'œil transforme la lumière reçue en influx nerveux qui sont transmis au cerveau. Mais de la même façon qu'un appareil photographique n'est pas seulement un capteur photonique, l'œil est constitué d'un ensemble sophistiqué de tissus venant guider la lumière vers la rétine. Dans son apparence externe et à l'œil nu, le globe oculaire se compose de la sclère (« blanc de l'œil ») et de la cornée. La sclère est une structure résistante, visant à protéger l'œil des dégâts mécaniques. Elle est recouverte (hormis sur sa partie antérieure) par la conjonctive, responsable de la production du mucus intégrant la composition du liquide lacrymal. Sur sa partie antérieure, la sclère est percée pour laisser place à la cornée. Si les yeux sont le miroir de l'âme, la cornée est la fenêtre des yeux. Il s'agit d'une membrane circulaire bombée et transparente constituée à 78% d'eau. Les larmes produites par les glandes lacrymales viennent l'alimenter en permanence et sont réparties sur toute sa surface par le battement des paupières. La cornée est la principale responsable de la focalisation des rayons lumineux pénétrant l'œil : à elle seule, elle assure près de 80% de la réfraction. Derrière la cornée se trouve une ouverture, nommée pupille, fonctionnant comme un diaphragme. La largeur de la pupille, contrôlée par les muscles lisses de l'iris, permet d'adapter la quantité de lumière pénétrant l'œil.

Pour assurer une mise au point aussi bien sur des objets lointains que proches, la focalisation doit être dynamique. Cette dynamique est rendue possible grâce au cristallin, petit disque fibreux situé à l'arrière de la cornée. La tension du cristallin est contrôlée par le corps ciliaire,

constitué d'un réseau de muscles (muscles ciliaires) et également responsable de la sécrétion de l'humeur aqueuse. L'ensemble cornée/cristallin joue le rôle de doublet optique dont la longueur focale peut être contrôlée grâce à la tension appliquée au cristallin. Postérieur au cristallin se trouve le corps vitré, où circulent les rayons lumineux avant de converger vers la surface tapissant l'intérieur du globe oculaire. Cette dernière, appelée rétine, est responsable de la conversion électromagnétique du signal lumineux en impulsions nerveuses. La Figure

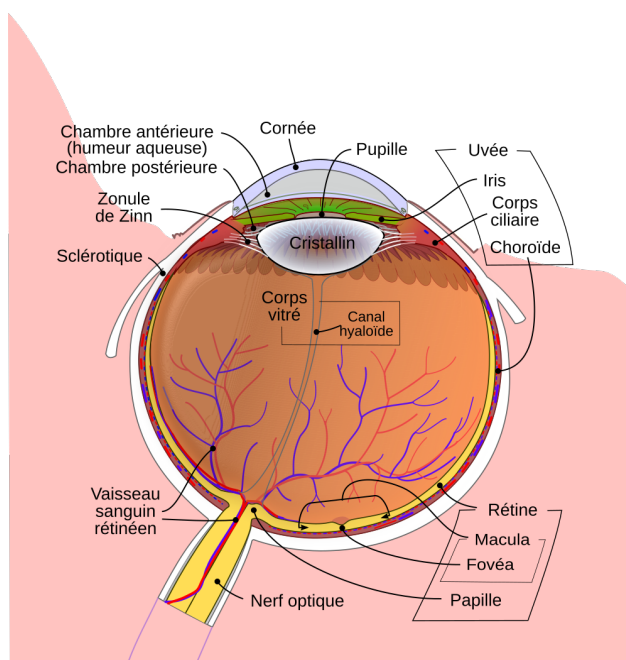


Figure 1.1 Schéma anatomique de l'œil humain. **Image libre de droits**, Wikipédia (2018).

1.1 synthétise en illustration cette anatomie de l'œil humain.

La rétine repose sur l'épithélium, lui-même rattaché à la choroïde qui est elle rattachée à la sclère. L'apport sanguin de la rétine provient majoritairement de la choroïde et dans une moindre mesure du réseau vasculaire rétinien. La rétine étant le siège de la transformation de l'onde électromagnétique en information nerveuse, l'activité neuronale y est particulièrement intense. Elle se fait par le biais d'une chaîne neuronale de transmission de l'information visuelle constituée de trois types de neurones :

- Les cellules photoréceptrices, éléments récepteurs, que l'on classe en deux catégories : les cônes et les bâtonnets. Les cônes sont responsables de la vision diurne et colorée. On en distingue trois types, en fonction de leur sensibilité spectrale (rouge, vert, bleu). Les bâtonnets sont chargés de la vision crépusculaire et nocturne, en « noir et blanc ».
- La cellule bipolaire, qui est un neurone de transmission, que l'on classe dans la catégorie des interneurons.

- La cellule ganglionnaire, qui est l'unité réceptrice fonctionnelle élémentaire. L'axone de la cellule ganglionnaire va former la fibre optique.

Cette chaîne neuronale constitue le corps de la rétine, que l'on divise en dix strates. Une brève description de celles-ci est proposée, suivant un ordre radial (depuis le corps vitré jusqu'à l'épithélium pigmentaire, c'est à dire le parcours d'un rayon lumineux).

1. L'interface entre le corps vitré et la rétine, appelée membrane limitante interne.
2. La couche des fibres optiques, qui contient les axones des cellules ganglionnaires, qui transmettent le signal visuel au cortex visuel. Les vaisseaux artériels et veineux sont situés dans l'épaisseur même de cette couche.
3. La couche des cellules ganglionnaires, qui contient le corps des cellules ganglionnaires.
4. La couche plexiforme interne, qui contient les cellules amacrines, qui sont une certaine catégorie de neurones (interneurones unipolaires).
5. La couche nucléaire interne, qui contient le corps cellulaire des cellules horizontales et des cellules bipolaires. Les cellules horizontales jouent un rôle dans la diffusion et la modulation des influx nerveux.
6. La couche plexiforme externe, qui contient les synapses entre les cellules bipolaires et les cellules photoréceptrices.
7. La couche nucléaire externe, elle est constituée des expansions internes des cellules photoréceptrices.
8. La membrane limitante externe
9. La couche des photorécepteurs. Cette couche, épaisse d'environ $40\mu m$ est constituée de 110 millions de bâtonnets et 7 millions de cônes, inégalement répartie dans la rétine.
10. L'épithélium pigmentaire. Celle-ci joue essentiellement le rôle d'écran, mais est également le siège d'échanges hydroélectriques et d'oxygènes, elle joue un rôle dans le métabolisme de la vitamine A et permet la régénération des photorécepteurs.

En plus de cette organisation en strates superposées, la rétine est également organisée en zones topographiques. La rétine centrale est composée de la fovéa, la macula et la papille. La macula est la région centrale de la rétine et en son centre se trouve la fovéa. La fovéa est le lieu de concentration des cônes, et donc par conséquent où l'acuité visuelle est à son maximum. La fovéola, au centre de la fovéa, est le lieu le plus mince de la région centrale de la rétine, du fait de l'absence de couches internes en cet endroit.

La papille représente la tête du nerf optique où convergent les fibres optiques. Cette zone de la rétine étant dépourvue de cellules photoréceptrices, il s'agit d'un point aveugle du champs de vision. De la papille émerge également un système d'artères intrarétiniennes, prenant en

charge l'apport en oxygène vers les couches internes de la rétine. Les couches externes (ainsi que les photorécepteurs) sont pour leur part alimentées par la choriocapillaire, en provenance de la choroïde.

La Figure 1.2 représente un fond d'œil sain. La forme circulaire centrale jaune-blanche est la papille, d'où sortent veines et artères (distinguables à leur couleur). La tache sombre et aux contours diffus sur la gauche est la région maculaire.



Figure 1.2 Exemple d'image de fond d'œil, issue de la base de données ROC (Niemeijer et al. (2010))

1.1.1 Pathologies rétiniennes

Les pathologies affectant la rétine peuvent aussi bien être des symptômes de maladies de l'œil ou systémiques. Cependant, les affections de la rétine sont nombreuses et en fournir une nosographie exhaustive n'est pas l'objet de ce mémoire. Une description des principales pathologies est proposée, les critères de sélection étant :

- La prévalence de la maladie (rapport du nombre de malades sur la population globale).
- Son incidence (nombre de nouveaux cas par an), indicateur qui motive grandement le développement d'algorithmes de dépistage automatique. Par exemple, au vu de l'incidence croissante de la rétinopathie diabétique, d'importants efforts sont consacrés à des algorithmes de dépistage pour maîtriser les flux anticipés de nouveaux patients.

- Les efforts déjà fournis par la communauté de traitement du signal et d’images pour une détection automatique de la maladie.

1.1.1.1 La Dégénérescence Maculaire Liée à l’Âge (DMLA)

La DMLA est la principale cause de pertes d’acuité visuelle chez les personnes de plus de 65 ans (Abràmoff et al. (2010)). Il en existe deux formes : l’atrophique (dite « sèche ») et l’exsudative (dite « humide »). La première est en général la plus fréquente, la moins dangereuse et se manifestant au premier stade de la maladie. Elle peut cependant évoluer sous la forme exsudative. Celle-ci, également appelée néo-vascularisation choroïdale, entraîne un risque important pour la vision. Elle se caractérise par une accumulation anormale de fluides dans et sous la rétine, potentiellement dangereuse si elle a lieu dans la région maculaire. Cliniquement, la DMLA peut se diagnostiquer par la présence de drusen dans la région maculaire. Ce sont des lésions jaunâtres, résultat d’une mauvaise élimination des pigments visuels (présents sur les photorécepteurs). En temps normal, ceux-ci se renouvellent en permanence, mais chez le patient atteint de DMLA, ils tendent à s’accumuler, ce qui peut avoir pour effet des tâches dans la vision du malade.

1.1.1.2 Le Glaucome

Le Glaucome est une neuropathie affectant le nerf optique. Elle est causée par une hypertension intra-oculaire venant compresser les fibres optiques. Au delà d’un certain seuil d’hypertension (on parle d’hypertonie oculaire), un blocage du transport axonal est provoqué, ce qui entraîne une dégénérescence de ces fibres. Cependant, le Glaucome peut également se manifester sans hypertonie oculaire, rendant son diagnostic compliqué. Auquel cas, le dépistage du Glaucome se fait grâce à des tests diagnostiques d’altération du champs visuel. Néanmoins, les tests de routine usuels ne permettent pas de le détecter. Une alternative est de dépister les prédispositions au Glaucome (tabac, hypertension artérielle, diabète, antécédents familiaux) (HAS (2006)).

L’évolution du Glaucome est généralement mesurée grâce au cup-to-disc ratio (CDR), qui mesure le rapport des largeurs de l’excavation et de la papille. L’augmentation de la pression intra-oculaire modifie la topologie de la papille avec pour effet un accroissement de ce rapport, ce qui permet donc de mesurer l’avancée du Glaucome.

1.1.1.3 La Rétinopathie Diabétique (RD)

La rétinopathie est une complication résultant du diabète sucré. Le diabète est causé par une défaillance des mécanismes de régulation de la glycémie qui entraîne une hyperglycémie. Or, un taux élevé de glucose dans le sang occasionne des dommages sur les petits et gros vaisseaux sanguins. La cible principale du glucose est l'endothélium (tissu constituant la paroi interne des vaisseaux, formée de cellules épithéliales) et la lame basale (sur laquelle repose les cellules épithéliales et chargée des apports nutritifs à l'endothélium). Un taux élevé de glucose et d'insuline dans le sang peut avoir pour conséquence de modifier les réactions des cellules de l'endothélium et d'augmenter la viscosité du sang, modifiant ainsi chimiquement certaines protéines. Dans le cas de la rétine, ces effets peuvent entraîner deux types de réactions :

- Une ischémie, que l'on définit comme la diminution de l'apport sanguin artériel à un organe, généralement du fait d'une obstruction de l'artère. Dans la rétine, cette obstruction peut conduire à la croissance de néo-vaisseaux pouvant entraîner des saignements et causer un détachement rétinien. A ce stade, la RD atteint le stade prolifératif.
- La rupture de la barrière hématorétinienne. Le liquide exsudé peut entraîner une enflure de la macula et endommager les photorécepteurs. Cette complication s'appelle un Œdème Maculaire Diabétique (OMD).

L'ischémie peut entraîner un blocage du transport axonal (déplacement de mitochondries, lipides, protéines et autres cellules dans le corps du neurone, permettant le renouvellement de celui-ci). L'accumulation axoplasmique dans la couche des fibres optiques entraîne des lésions appelées Cotton wool spots, d'apparence blanche et aux frontières floues (McLeod (1981)). Mais la principale cause de cécité au sein de la population atteinte de diabète est une conséquence d'un OMD. Cliniquement, il s'observe principalement par l'épaississement de la rétine dans la région maculaire et par la présence d'exsudats durs proches de la macula. Les exsudats durs sont des résidus de lipides qui se déposent dans la couche plexiforme externe.

Parmi les différentes maladies présentées, une place prépondérante est accordée à la RD dans la communauté d'analyse et du traitement d'images. Pour l'expliquer, il faut tout d'abord rappeler son impact sociétal :

- La prévalence de la maladie est particulièrement importante. Près de 34 % des individus diabétiques sont atteints de RD, soit approximativement 93 millions de malades à travers le monde (Yau et al. (2012)). Parmi eux, 28 millions sont atteints de RD menaçant leur vision à court terme.
- Son incidence est en forte croissance, notamment en Asie, ce qui laisse à craindre une saturation des services de soin ophtalmiques (Zheng et al. (2012)).

Or, à son stade non-prolifératif, la RD peut être traitée au laser. Ce dernier va venir détruire

les néo-vaisseaux qui risqueraient d'entraîner un décollement de la rétine. Cette opération est appelée photocoagulation panrétinienne. Au stade prolifératif, le traitement est nettement plus complexe et requiert de remplacer le corps vitré à l'intérieur de l'oeil (vitrectomie) ; cependant, cette opération ne permet pas de récupérer l'acuité visuelle originelle. De fait, il est donc important de déceler et suivre régulièrement la progression de la maladie. Ayant commencé à porter ses fruits (Khawaja et al. (2017)), l'automatisation du processus de dépistage porte la promesse d'une réduction des coûts, d'une détection accélérée et d'un suivi amélioré de la progression de la maladie. La *Early Treatment Diabetic Retinopathy Study (ETDRS)* (Early Treatment Diabetic Retinopathy Study Research Group et al. (1991)) définit une nomenclature de l'avancée de la maladie en fonction des lésions observées. En résumé, la maladie peut se classer en trois types selon son stade :

1. non-proliférative (légère, modérée, sévère). Les lésions observées sont des hémorragies, des micro-anévrismes et/ou des exsudats durs. La tortuosité des vaisseaux vasculaires peut également être un symptôme observable.
2. proliférative (précoce, haut-risque, avancée). On observe des néo-vascularisations, y compris dans l'iris, des hémorragies dans le corps vitré et/ou un décollement de la rétine.
3. OMD. La présence d'un œdème n'est pas forcément cliniquement significative. Elle le devient si l'épaississement de la rétine et/ou l'apparition d'exsudats durs se font à moins de $500\mu m$ de la fovéa.

1.1.1.4 Autres maladies

Les trois maladies préalablement décrites sont les principales responsables des pertes progressives de la vision. Néanmoins, elles sont loin d'être les seules affections touchant la rétine. La rétinopathie hypertensive et l'athérosclérose entraînent par exemple un changement du rapport artère/veine (AVR) dans la rétine. Ceci peut avoir des conséquences graves ; par exemple, une diminution de ce rapport augmente le risque d'infarctus du myocarde. La rétinopathie hypertensive est la conséquence d'une hypertension artérielle, l'athérosclérose est une manifestation affectant les artères par l'apparition de plaques d'athérome (dépôts de graisse à la frontière de l'endothélium, lipides, calcium et cellules sanguines, entraînant une diminution du débit artériel). L'hypertension est également une cause potentielle d'ischémie et donc d'apparition de Cotton wool spots. D'autres maladies systémiques peuvent également se déceler dans la rétine. Un gonflement de la papille (œdème papillaire) peut par exemple être du à une augmentation de la pression intracrânienne, une tumeur cérébrale, une insuffisance respiratoire, un syndrome de Guillain-Barré... Par ailleurs, des maladies vasculaires

systémiques peuvent également déclencher des occlusions artérielles et veineuses (blocage soudain du vaisseau, souvent du fait d'un caillot).

1.2 Imagerie diagnostique de la rétine

Le diagnostic des différentes maladies décrites ne peut s'obtenir sans observation de la structure interne de l'œil. L'invention du premier ophtalmoscope est attribuée à von Helmholtz au milieu du XIX^e siècle. Il devint très rapidement un instrument incontournable pour les ophtalmologistes par la possibilité qu'il offre d'observer la rétine de façon complètement non invasive. Il fut grandement amélioré au cours du XX^e siècle, notamment pour corriger la focalisation chez les patients atteints de myopie. En 1910, Allvar Gullstrand proposa de photographier l'image formée par l'ophtalmoscope : la caméra fond d'œil était née. Il obtint d'ailleurs le prix Nobel de médecine en 1911 pour ses travaux sur la dioptrie de l'œil.

L'imagerie de fond d'œil est définie par Abràmoff et al. (2010) comme étant :

Le processus par lequel on obtient, grâce à la lumière réfléchie, une représentation en deux dimensions des tissus rétinien semi-transparents tri-dimensionnels projetés sur le plan de l'image.¹

Le défi de taille de l'acquisition est l'illumination de la rétine, notamment du fait de l'étroitesse de la pupille. En général, un collyre mydriatique est utilisé pour provoquer sa dilatation. Cependant, avec les progrès technologiques récents et en fonction des modalités, le rétinographe peut se faire sans agent mydriatique. Une autre difficulté provient du fait de la réflexion spéculaire sur les interfaces de la cornée et le cristallin. Ces réflexions imposent aux faisceaux incident et réfléchi (sur la rétine) de prendre deux chemins séparés à travers la pupille, ce qui réduit d'autant plus l'ouverture effective de celle-ci.

Selon la définition de Abràmoff et al. (2010), plusieurs modalités peuvent être appelées fond d'œil, en plus de la simple capture photographique de l'image formée par l'ophtalmoscope.

La photographie fond d'œil stéréo En capturant deux images prises selon deux angles de vues différents et en les calibrant, il est possible d'obtenir une approximation de la profondeur. On obtient donc une représentation en relief de la surface rétinienne.

L'imagerie de fond d'œil hyperspectrale Au lieu d'illuminer la rétine à la lumière blanche, on vient successivement envoyer des rayons à des longueurs d'onde croissantes. On

1. « *We define fundus imaging as the process whereby a 2-D representation of the 3-D retinal semi-transparent tissues projected onto the imaging plane is obtained using reflected light.* », Abràmoff et al. (2010)

mesure la quantité de lumière réfléchie par longueur d'onde pour obtenir la réponse spectrale de la rétine. Cette modalité, récente, reste encore peu utilisée. Kaluzny et al. (2017) propose son usage pour l'estimation d'oxymétrie rétinienne (permettant ainsi de distinguer veines et artères) et faire des mesures de densité de pigment optique dans la région maculaire.

L'ophtalmoscopie à balayage laser (*Scanning Laser Ophthalmoscopy (SLO)*) Au lieu d'illuminer globalement la rétine, un mince faisceau laser balaye la surface. L'acquisition se fait donc progressivement. Cette technique, développée au début des années 1980, présente l'avantage d'une haute résolution spatiale. Par ailleurs, elle limite les risques d'éblouissements ou de phototraumatismes. Cette modalité ne nécessite pas de dilater la pupille par application d'un agent mydriatique.

L'angiographie en fluorescence A la différence du simple fond d'oeil, un agent de contraste (fluorescéine ou indocyanine) est injecté au patient. Il se retrouve dans la circulation sanguine et la quantité de photons qu'il émet est mesurée. La façon dont l'agent se répand ainsi que la manière dont le signal s'évanouit permet de caractériser le réseau vasculaire et faire ressortir certaines lésions (tels les micro-anévrysmes).

La caméra fond d'œil présente cependant deux inconvénients. D'une part, elle est la représentation bidimensionnelle d'un tissu tridimensionnel. D'autre part, elle ne permet d'imager que la surface de la rétine : il est donc impossible d'observer la choroïde par exemple. Ces deux problématiques ont été résolues grâce à l'invention de l'Tomographie en Cohérence Optique (OCT) par Huang et al. (1991) permettant d'obtenir une représentation tridimensionnelle de la rétine et des couches inférieures. L'imagerie OCT repose sur le principe de l'interféromètre de Michelson. Sans entrer dans les détails théoriques du fonctionnement de l'appareil, son principe fondamental repose sur la division d'un même faisceau lumineux (en général proche de l'infrarouge) pour guider les deux sous-faisceaux dans les deux bras de l'appareil. Le premier bras constitue la référence, le second guide la lumière vers le fond d'oeil. Les rayons réfléchis (par la référence d'une part, par une des interfaces rétinienne d'autre part) sont placés en interférence. En déplaçant la référence le long du bras et en analysant les franges d'interférences, on peut estimer la différence de marche entre les deux rayons réfléchis et obtenir ainsi une estimation de la profondeur de l'interface rétinienne sur laquelle a eu lieu la réflexion. En balayant longitudinalement et transversalement la surface de la rétine, on peut reconstituer sa structure en trois dimensions.

En pratique, ce principe de fonctionnement basique est peu utilisé. En effet, le déplacement mécanique de la référence est une source d'imprécision, en plus de rallonger le temps d'acquisition. Déployée depuis peu, l'OCT dans le domaine fréquentiel procure un signal moins

bruité et obtenu plus rapidement.

L'OCT est une technique d'imagerie au champs d'application plus élargi que le fond d'œil, car elle peut également être déployée en cardiologie pour imager l'intérieur des artères coronaires via une sonde placée sur un cathéter. En ophtalmologie, elle permet de confirmer un diagnostic réalisé au fond d'œil et de détecter le Glaucome (en mesurant l'épaisseur des différentes couches). La DMLA peut également être détectée en analysant l'épaisseur et la rigidité de la choroïde. Par ailleurs, l'OCT peut être combinée avec des méthodes d'angiographies, permettant une étude affinée des microvasculatures de la rétine. La Figure 1.3 est une coupe OCT non aplanie. En général, un algorithme de prétraitement est appliqué pour corriger la sphéricité de la rétine, ce qui facilite la segmentation des interfaces. Sur cette image, celles-ci se distinguent nettement. La surface large et non homogène que l'on voit sous la rétine est la choroïde.

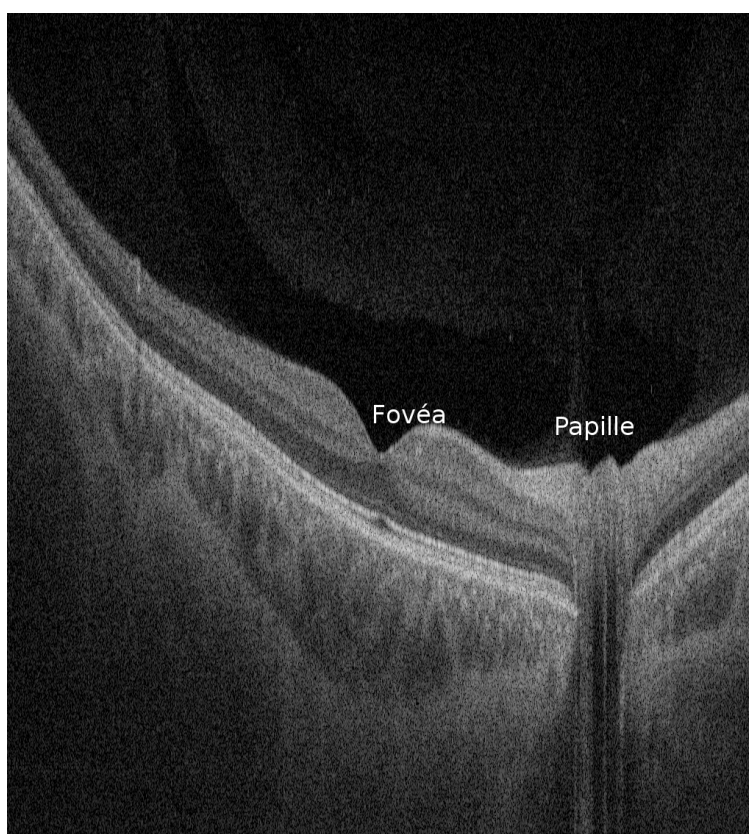


Figure 1.3 Coupe OCT de la rétine. On y observe la forme sphérique de la surface de la rétine et l'emplacement de la fovéa (lieu d'épaisseur minimale de la rétine) et de la papille (d'où arrive le nerf optique).

1.3 Objectifs généraux

Cet aperçu des enjeux cliniques et des modalités d'imagerie permet de contextualiser les travaux de recherche menés par la communauté de traitement du signal et les enjeux de santé publique qui y sont liés. Le chapitre qui suit en présentera les principaux tenants, l'objectif commun à tous étant de progresser dans l'élaboration d'un système de diagnostic automatisé.

Dans le cadre de cette maîtrise, les travaux de recherche se sont concentrés sur l'imagerie de fond d'œil, sans cibler spécifiquement une maladie en particulier, mais en privilégiant la détection et la segmentation de lésions, celles-ci permettant par la suite de remonter à un diagnostic.

1.4 Plan du mémoire

Ce mémoire est structuré en cinq chapitres. Dans le premier chapitre, un état de l'art des travaux portant sur l'analyse automatisée des images de fond d'œil est présenté. Cette revue de littérature est structurée en sections qui correspondent aux différentes briques séquentielles composant la chaîne de traitement d'un système de diagnostic automatique : évaluation de la qualité de l'acquisition, prétraitement, méthodes de segmentation et classification. Ce mémoire reposant sur l'apprentissage profond, une section est consacrée aux évolutions de ce domaine en terme de segmentation. Enfin, une section est également réservée à la description des bases de données publiques disponibles, les données étant la variable essentielle des travaux de recherche menés au long de cette maîtrise.

La seconde partie synthétise les deux précédentes pour dresser la problématique de recherche ayant guidé cette maîtrise, en précisant les objectifs spécifiques visés.

Le cœur du mémoire repose sur le chapitre trois. Il s'agit d'un article soumis en Juillet 2018 à la revue *IEEE Transactions on Medical Imaging* portant sur la segmentation des lésions de la rétine. La méthodologie repose sur une nouvelle architecture de réseau de neurones convolutive multitâche, entraînée selon une méthode inédite permettant de généraliser l'apprentissage sur des bases de données comportant peu d'annotations cliniques. L'article met l'accent sur la validation des résultats obtenus et une discussion des avantages et limitations est proposée.

Le quatrième chapitre de ce mémoire présente les travaux parallèles menés pour le développement d'une plateforme clinique permettant de faciliter l'annotation des pathologies par les médecins. Les objectifs de développement ainsi que les choix technologiques réalisés y sont présentés.

Le dernier chapitre propose une discussion synthétique des différents travaux réalisés, pointant les limitations du modèle actuel et proposant des recommandations sur les travaux à conduire dans le futur.

CHAPITRE 2 REVUE DE LITTÉRATURE

Dans ce chapitre, nous allons passer en revue différents pans de la littérature concernant la détection automatique de pathologies dans le fond d'œil. Pour ce type d'images, la chaîne de traitement de l'acquisition suit généralement les mêmes étapes quelque soit la méthodologie. En premier lieu, une fois l'acquisition numérique obtenue, une évaluation de la qualité de l'image est faite. Cette évaluation peut être faite directement par le clinicien, qui décidera dès lors si l'acquisition doit être refaite ou peut être conservée. Le cas échéant, l'image est ensuite stockée pour utilisation ultérieure. L'évaluation de la qualité peut aussi être le fait d'un algorithme, comme présenté dans la section 2.1. L'étape suivante est en général un pré-traitement appliqué à l'image pour en rehausser les contrastes et accentuer le rapport signal sur bruit (Signal-to-noise Ratio (SNR)). A partir de là, à des fins de diagnostic automatique, trois stratégies possibles existent. La première va venir tenter d'identifier individuellement la présence de biomarqueurs anormaux, soit par simple détection (localisation de zones de l'image contenant potentiellement une anomalie) soit par segmentation (recherche des frontières précises du biomarqueur). La seconde stratégie est une approche globale au niveau de l'image. Des caractéristiques extraites au niveau de l'image sont utilisées pour une classification directe à des fins de dépistage (image saine ou malade). Certains systèmes, en plus du simple dépistage, peuvent également associer un stade d'avancement de la maladie (*grading*). La troisième approche est une méthode hybride entre les deux premières, les marqueurs identifiés par la première approche étant utilisés comme vecteurs caractéristiques classifiés via la deuxième approche. Chacune de ces trois approches repose sur le même schéma : extraction de caractéristiques (localement pour la première, globalement pour les deux suivantes) et classification selon ces caractéristiques. La Figure 2.1 résume l'ensemble de cette chaîne. Dans le monde de la recherche et avant un éventuel déploiement clinique, une dernière étape est ajoutée à cette chaîne : il s'agit de l'évaluation de la performance des algorithmes, par des mesures de différentes métriques.



Figure 2.1 Chaîne de traitement automatique d'une image de fond d'œil.

La revue de littérature proposée suit l'ordre des différentes étapes de cette chaîne.

2.1 Evaluation de la qualité des acquisitions

L'étude dirigée par Philip et al. (2005) pour évaluer l'efficacité d'un centre de dépistage à distance (le diagnostic n'est pas fait en même temps que l'acquisition), a mis en évidence que 11.9% des patients (sur 5575) ont du être rappelés pour procéder à une nouvelle acquisition du fait de la mauvaise qualité d'image obtenue à la première tentative. Par ailleurs, parmi les images acceptées, 15.5% d'entre elles sont considérées comme seulement partiellement exploitables. Plusieurs raisons sont mises en avant pour expliquer ce taux élevé (mais similaire à celui d'études menées dans d'autres pays) : pendant la période de l'étude, deux nouveaux praticiens encore en phase d'apprentissage ont pu fausser les statistiques. Par ailleurs les caméras haute-résolution, du fait du flash intensif qu'elles utilisent, peuvent causer des reflets indésirables. Enfin, l'étude démontre que les patients ayant subi une opération de la cataracte sont susceptibles d'être sujets à une photographie de moins bonne qualité.

Évaluer la qualité d'une image en amont de la chaîne est donc nécessaire, à la fois pour réduire les coûts financiers et pour accélérer le dépistage. Cette évaluation peut se faire suivant plusieurs critères. Une manière très générique d'évaluer la qualité d'une image est de construire une métrique sur l'illumination, le contraste et la finesse des détails. Une telle approche semblable est explorée par Herman et al. (2009) qui se base localement sur la finesse et l'illumination et les combine en un score unique. La finesse (*sharpness*) est calculée dans le domaine spectral comme la moyenne des amplitudes des fréquences présentes dans un certain intervalle (après application d'un filtre passe-bande). L'illumination est définie comme la somme de la luminosité (*brightness*) et du contraste, obtenus respectivement via l'intensité moyenne et l'écart-type local. Plus récemment, Wang et al. (2016) a proposé une autre approche basée sur la modélisation du Human Visual System (HSV) (Système visuel humain). Pour cela, trois caractéristiques sont extraites de l'image : l'appréhension multi-canal (*Multi-Channel Sensation*), qui correspond à la perception de l'illumination et de la couleur, le « flou à peine perceptible » (*Just Noticeable Blur*) se basant sur la capacité à extraire des vaisseaux nets dans l'image et enfin la fonction de sensibilité au contraste (*Contrast Sensitivity Function* (CSF)). Cette dernière caractéristique est calculée sur une version filtrée de l'image par un filtre passe-bande. L'image filtrée est notée \hat{I} et le contraste est calculé en se basant sur la dispersion de son histogramme :

$$f = \frac{Q_3(\hat{I}) - Q_1(\hat{I})}{\max(\hat{I}) - \min(\hat{I})} \quad (2.1)$$

où $Q_{1,3}$ représentent respectivement le premier et troisième quartile de la distribution. Le résultat final est obtenu à partir des trois caractéristiques HSV, utilisées pour entraîner un Support Vector Machine (SVM) (machine à vecteurs de supports) à prédire la qualité globale de l'image.

Ces approches génériques, bien qu'efficaces en pratique, ont l'inconvénient de ne pas s'appuyer sur l'identification des marqueurs attendus dans le fond d'œil : n'étant pas construites par rapport aux structures anatomiques de la rétine, elle ne garantissent pas que celles-ci soient observables en pratique. Par ailleurs, elles ne permettent généralement pas de distinguer les marqueurs biologiques des tâches ou reflets liés à l'acquisition. A contre-pied de ces approches, des méthodes structurelles ont été explorées, prenant en compte les tissus qu'on s'attendrait à observer dans un fond d'œil de bonne qualité. Par exemple, Usher et al. (2003) propose un score de qualité se basant simplement sur la quantité de vaisseaux segmentés. Plus élaborée, Fleming et al. (2006) propose une méthode se basant sur la clarté dans la région maculaire et sur le champs de visibilité dans la région entourant le disque optique. La macula est obtenue en cherchant le point de corrélation maximale avec un modèle de fovéa paramétré et le disque optique par transformation de Hough. Au niveau de la macula, la clarté est obtenue comme Usher et al. (2003), c'est à dire par segmentation du réseau vasculaire et en additionnant les pixels appartenant à des vaisseaux. Plus la clarté est haute, plus les vaisseaux dans la région maculaire sont visibles, plus l'image est considérée de bonne qualité. Pour le champ de visibilité, il est défini comme un agrégat d'un ensemble de contraintes. Par exemple, la distance entre le centre du disque optique et le bord de l'image le plus proche doit être supérieure au rayon d'un disque optique moyen (valeur empirique), la macula doit se situer à deux diamètres de disque optique du bord de l'image, etc...

2.2 Méthodes de segmentation

Pour chacune des structures anatomiques considérées, de très nombreuses méthodes de segmentation ont déjà été proposées, il serait illusoire de vouloir en faire une revue exhaustive. Un choix a été fait pour fournir une description des principales approches, celles proposant des fondements théoriques solides et ayant influencé un grand nombre de travaux ultérieurs. En particulier, pour les marqueurs sains, l'accent est mis sur les méthodes paramétriques (c'est à dire sans apprentissage). En effet, la détection des marqueurs sains est en général un préalable à la détection des marqueurs pathologiques et en tant que tel doit pouvoir se généraliser à n'importe quelle base sans ré-entraînement (les annotations des cliniciens n'étant pas forcément disponibles). Pour autant (et sans grande surprise), la plupart des publications récentes font état de performances supérieures en utilisant des réseaux de neurones.

La revue de littérature sur la segmentation des lésions est couverte dans la section 4.1, dans l'article joint à ce mémoire, qui se focalise sur cet aspect là.

2.2.1 Segmentation des marqueurs normaux

La segmentation des structures anatomiques est souvent employée en préalable de méthodes de détection des lésions, permettant d'éviter de compter ces marqueurs en tant que faux positifs. Par exemple, les petits vaisseaux peuvent être aisément confondus avec des hémorragies, le disque optique avec des lésions claires (drusen, exsudats, Cotton wool spot). Par conséquent, en préalable de la détection de lésions, les principaux marqueurs segmentés sont le disque optique, la macula et le réseau vasculaire. Cependant, la segmentation de ces trois tissus est en elle-même nécessaire à certaines fins diagnostics. A titre d'exemple, la segmentation du disque optique permet de mesurer le CDR et détecter un éventuel Glaucome (voir section 1.1.1.2) tandis que la segmentation sémantique du réseau vasculaire (c'est à dire segmentation et distinction entre veines et artères) permet de calculer l'AVR (section 1.1.1), mesure de rétinopathie hypertensive.

Disque Optique (DO) et macula La segmentation du DO est généralement réalisée en préalable de la segmentation de l'arbre vasculaire ou des lésions claires. Dans le premier cas, il s'agit de trouver le point de départ des vaisseaux, dans le second cas, d'éliminer une potentielle source de faux positifs. La segmentation est parfois faite conjointement avec la détection de la macula. En effet, la distance relativement constante entre ces deux marqueurs permet d'obtenir un invariant d'échelle. Il arrive parfois que la démarche inverse soit mise en œuvre, les vaisseaux étant d'abord segmentés pour identifier la position du DO. La segmentation du DO s'accompagne parfois de la segmentation de l'excavation papillaire pour la mesure du CDR (voir Figure 2.2).

La difficulté de la segmentation du disque optique réside dans les grandes variations de forme, taille et couleur existantes. Par ailleurs, ses frontières peuvent être floues, de contrastes non-uniformes et non continues (à cause de la sortie des vaisseaux sanguins). De ce fait, deux catégories de travaux ont été menées sur le disque optique. La première se contente de trouver la localisation de son centre, la seconde va plus loin et met l'accent sur la segmentation. Pour la localisation, les travaux exploitent généralement la structure vasculaire émergeant du disque optique. Ainsi, pour identifier le centre du DO, Sinthanayothin et al. (1999) analyse le lieu de variance maximale (due à la structure vasculaire), Foracchia et al. (2004) recherche le point d'intersection des quatre principaux vaisseaux et Youssif et al. (2008) se base sur la direction de convergence des vaisseaux.

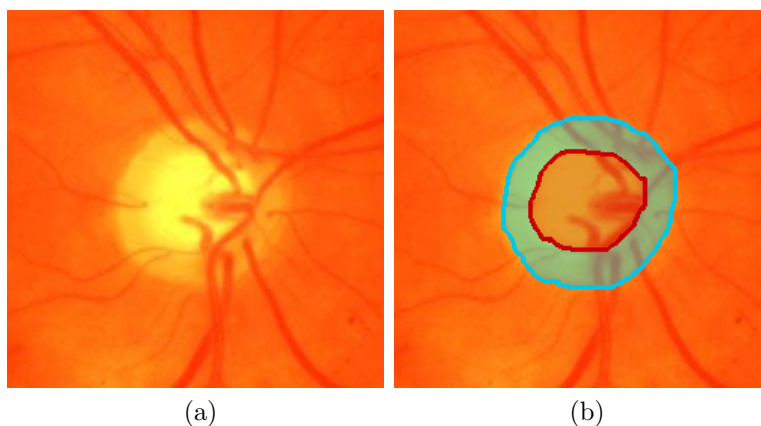


Figure 2.2 Segmentation du disque optique et de l'excavation papillaire.

En terme de segmentation, la plupart des approches repose sur des méthodes basées sur un modèle (*template-based methods*), qui approximent les frontières du DO. Plus éparés, des travaux se basent également sur des méthodes basées sur la recherche active de contours. Dans le premier cas, le modèle est généralement un cercle ou une ellipse paramétrés. Cette approche est explorée par Lalonde et al. (2001) qui s'appuie sur la distance de Hausdorff pour évaluer la proximité entre des cercles de différents rayons (les modèles) et la carte des contours de différentes régions candidates. Dans cette même optique, Wong et al. (2008) propose d'ajuster une forme elliptique paramétrée aux contours du DO. La localisation du DO est obtenue par analyse de l'histogramme et les contours initiaux sont obtenus simplement par seuillage. L'avantage des techniques se basant une formée paramétrée (cercle ou ellipse) est d'être relativement robuste aux contours non continus du DO, les discontinuités étant causées par le réseau vasculaire qui les traverse.

Aquino et al. (2010) propose également une forme d'ajustement de modèle, en utilisant une transformation de Hough. La première étape est la localisation de pixels appartenant au DO. Pour ce faire, un système de votes entre trois méthodes est mis en place. La première méthode recherche le lieu de contraste maximum (là où la différence d'intensité est la plus forte, les vaisseaux sombres contrastant avec le DO claire), la seconde se base sur le lieu de variance maximale et un seuillage selon la méthode d'Otsu (pour déterminer si le lieu de variance maximale correspond à une tâche sombre ou claire) et la dernière méthode repose sur un filtrage des hautes fréquences par un filtre passe-bas. Une fois la localisation faite, les vaisseaux sont retirés du disque et les contours approximatifs de celui-ci sont extraits en utilisant des filtres de Prewitt. Une transformation de Hough circulaire permet d'obtenir un disque bien circulaire.

Cependant, ces quelques exemples ne suffisent pas à couvrir la variété des méthodes exis-

tantes : Walter and Klein (2001) utilise l’algorithme Watershed, Abramoff et al. (2007) propose d’utiliser les k-plus proches voisins pour classifier chaque pixel en fonction d’un ensemble de caractéristiques extraites par pixel, Morales et al. (2013) s’appuie sur des opérations morphologiques en amont de l’algorithme Watershed pour extraire le disque optique. A des fins de détection uniquement, Rahebi and Hardalaç (2016) propose même une approche simulant le comportement des libellules évoluant en essaim vers les sources de lumière ! Sur plusieurs métriques (voir section 2.4.3), les meilleurs résultats de segmentation sont revendiqués par Al-Bander et al. (2018), utilisant des réseaux de neurones convolutifs suivant une architecture U-Net.

Concernant la région maculaire, la littérature est moins dense mais repose généralement sur les mêmes approches. A titre d’exemple, Taori et al. (2016) s’appuie sur des opérations morphologiques pour extraire les caractéristiques utilisées dans un algorithme de croissance de régions. La position approximative de la macula (et donc le point de départ de l’algorithme) est obtenue par analyse d’histogramme. Onal et al. (2016) propose une approche dynamique de classifications des contours après extraction des lésions sombres susceptibles d’être confondues. La zone d’intérêt est identifiée par seuillage simple. Elle est ensuite projetée dans l’espace polaire : la macula étant de forme circulaire, dans cet espace, elle devient approximativement une droite. Une méthode par contours actifs est ensuite utilisée pour trouver les frontières de la macula dans cet espace.

Réseau vasculaire La segmentation du réseau vasculaire a fait l’objet de nombreux travaux, se basant sur des opérateurs de seuillage, de calcul de gradient, morphologiques ou en encore en utilisant des méthodes d’apprentissages machine. Une des premières étapes communes à ces méthodes est d’appliquer un pré-traitement permettant de faire ressortir les structures typiquement longitudinales des vaisseaux. La méthode la plus commune repose sur les travaux de Frangi et al. (1998). A l’origine, il s’agissait d’exploiter la forme tubulaire des vaisseaux en trois dimensions, mais son application à l’image 2D est directe. Pour chaque pixel de l’image \mathbf{I} , on calcule la Hessienne définie par :

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 \mathbf{I}}{\partial x^2} & \frac{\partial^2 \mathbf{I}}{\partial x \partial y} \\ \frac{\partial^2 \mathbf{I}}{\partial y \partial x} & \frac{\partial^2 \mathbf{I}}{\partial y^2} \end{pmatrix} \quad (2.2)$$

Dans ce contexte, la différenciation est estimée par convolution avec la dérivée de la fonction gaussienne :

$$\frac{\partial}{\partial x} \mathbf{I}(\mathbf{x}, \sigma) = \sigma \cdot \mathbf{I} * \frac{\partial}{\partial x} \mathbf{G}(\mathbf{x}, \sigma) \quad (2.3)$$

où la Gaussienne est définie en dimension 2 par :

$$\mathbf{G}(\mathbf{x}, \sigma) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right) \quad (2.4)$$

La convolution par la dérivée seconde de la Gaussienne par rapport à l'axe x ou y mesure le contraste entre la région située dans l'intervalle $(-\sigma, \sigma)$ et à son extérieur (selon l'axe x ou y considéré). De fait, la Hessienne permet donc de caractériser le contraste dans cet intervalle selon chacune des directions (x , y et xy). Pour connaître la direction privilégiée, la décomposition de la Hessienne en vecteurs propres est opérée. Ainsi, on peut obtenir la direction de courbure au pixel considéré. Si le rapport des valeurs propres est proche de 1, le pixel ne se situe pas le long d'une courbe, mais plutôt sur une structure sphérique. Inversement, si ce rapport est grand, le pixel se situe sur une forme allongée. Pour faire ressortir ces dernières et discriminer les pixels dont les deux valeurs propres sont similaires, Frangi et al. (1998) propose d'utiliser la mesure de *vesselness* :

$$V(\sigma) = \exp\left(-\frac{\lambda_1(\sigma)^2}{\lambda_2(\sigma)^2}\right)(1 - \exp(-(\lambda_1(\sigma)^2 + \lambda_2(\sigma)^2))) \quad (2.5)$$

où $\lambda_1(\sigma)$, $\lambda_2(\sigma)$ sont les valeurs propres (par ordre croissant) de la Hessienne calculée à une échelle σ donnée. En faisant varier la valeur de σ , on obtient une approche multi-échelle.

L'approche de Frangi et al. (1998) n'est cependant pas optimale dans plusieurs cas de figures. Deux vaisseaux parallèles et proches auront tendance à être réunis si la résolution spatiale employée est trop grossière. Inversement, l'effet *Copper Wire*, dû à une réflexion plus prononcée de la lumière sur le milieu des principales artères (voir Figure 2.3), perturbe le profil gaussien du vaisseau. La méthode de Frangi et al. (1998) peut en être affectée ce qui entraîne une double segmentation au lieu d'une simple. De nombreuses autres approches, se basant sur le principe de filtre adaptatif, ont également été proposées. Ces méthodes font nécessairement l'hypothèse du profil Gaussien des vaisseaux et par conséquent se laissent également tromper par l'effet *Copper Wire*. D'autres travaux proposent d'utiliser des méthodes de classification. Néanmoins, celles-ci nécessitent des références annotées manuellement pour l'apprentissage (uniquement disponible en quantité limitée) et ont tendance à mal généraliser sur des bases de données inconnues. Partant de ces constats, Nguyen et al. (2013) propose une approche entièrement paramétrique devenue une référence nommée *Multi-Scale Line Detector (MSLD)* (détection de lignes multi-échelle). Toutes les opérations s'effectuent au sein d'une fenêtre glissante sur l'image. Au sein de cette fenêtre, on calcule la réponse linéaire du pixel central. Pour cela, on mesure la moyenne de l'intensité des pixels le long de douze segments orientés selon une direction propre. Chaque segment est orienté à 15° du précédent (cet angle définis-

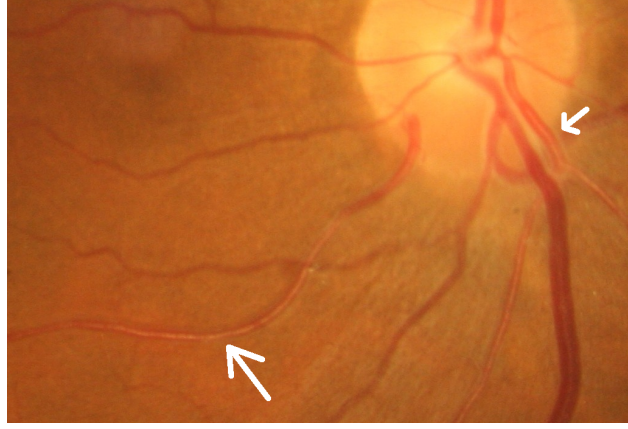


Figure 2.3 Effet *Copper Wire* sur des artères d'un patient atteint de rétinopathie hypertensive
Image libre de droit, Wikipedia (2018)

sant la résolution angulaire), pour couvrir l'ensemble des orientations possibles. La réponse linéaire est définie comme la différence entre la moyenne du profil d'intensité maximale et l'intensité moyenne dans la fenêtre (moyenne des douze profils). Plus formellement, en notant R la réponse linéaire du pixel, $I_j^W, j \in \{1, \dots, 12\}$ l'intensité moyenne le long de chacun des profils j de taille W alors :

$$R = \max_j I_j^W - \frac{1}{12} \sum_j I_j^W \quad (2.6)$$

La réponse sera bien maximale pour un pixel se trouvant sur une structure allongée. Pour un pixel se trouvant sur une forme sphérique (type hémorragie) ou sur du fond, la moyenne des profils sera proche du profil d'intensité maximale. Par conséquent, la réponse linéaire sera faible. La Figure 2.4 donne un exemple de réponses linéaires sur une image de fond d'œil faisant bien ressortir la structure linéaire des vaisseaux.

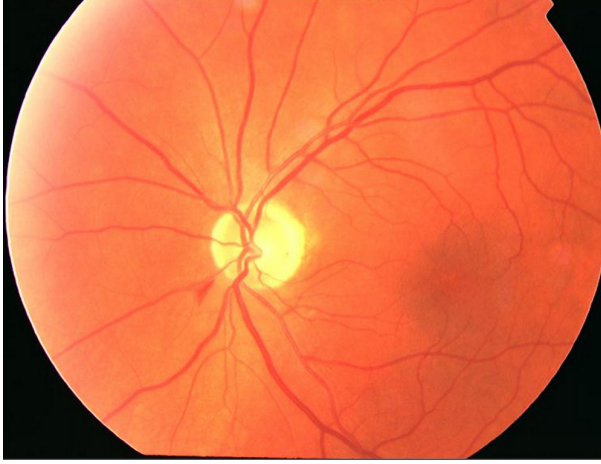
L'équation 2.6 n'est pas multi-échelle, car elle se base sur une évaluation fixée de l'épaisseur a priori d'un vaisseau. Si le segment selon lequel on mesure le profil d'intensité est plus court que l'épaisseur de ce faisceau, la réponse linéaire sera faible (la moyenne des profils étant égale au profil maximal). Inversement, si ce segment est trop long, le risque est de joindre deux vaisseaux parallèles proches. Nguyen et al. (2013) propose par conséquent d'évaluer la réponse linéaire à différentes échelles, c'est à dire pour différentes longueurs L de segments :

$$R^L = \max_j I_j^L - \frac{1}{12} \sum_j I_j^L \quad (2.7)$$

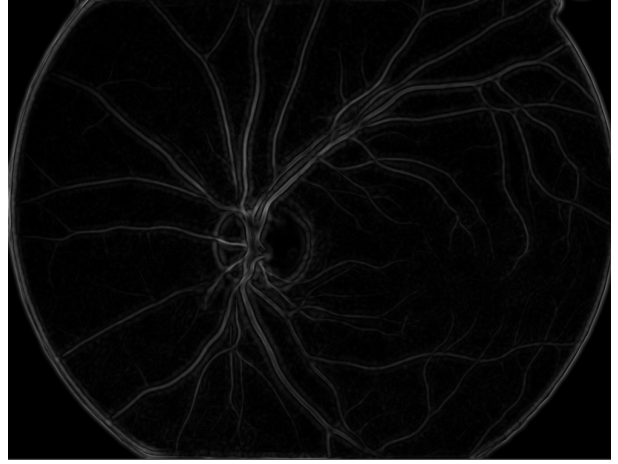
La réponse multi-échelle n'étant qu'une combinaison linéaire des réponses à chacune des échelles :

$$R_{multi-échelle} = \frac{1}{n_L + 1} \left(\sum_L R^L + I_g^{(inv)} \right) \quad (2.8)$$

où $I_g^{(inv)}$ est le canal vert inversé de l'image originale.



(a) Image originale



(b) Réponse linéaire à l'algorithme de détection de lignes pour $W = 15$.

Figure 2.4 Démonstration de la détection de ligne selon la méthode de Nguyen et al. (2013)

2.3 Approche globale de la classification et de la gradation de la maladie

Dans les sections qui précèdent, les travaux se concentrent sur la recherche des structures anatomiques au niveau du pixel (segmentation et/ou détection). A partir des sorties de l'algorithme, il est cependant possible de dériver des métriques sur l'existence d'une potentielle pathologie au niveau global de l'image. Cette approche est ascendante : des caractéristiques mesurées au niveau local permettent de dresser un bilan global. De plus, les caractéristiques locales ne sont pas choisies par hasard : elles correspondent à des symptômes documentés cliniquement. A ce titre, le fait même de chercher telle ou telle structure induit une forme de biais visant à rapprocher l'algorithme du mode de fonctionnement du clinicien.

A contrario, l'approche globale cherche à se défaire de ce biais. Intuitivement, l'idée se résume à laisser à l'algorithme le choix des caractéristiques permettant de discriminer les images en fonction de leur diagnostic. Par définition, ces approches sont le propre de l'apprentissage machine.

Pires et al. (2017) propose une méthode basée sur la classification de caractéristiques extraites automatiquement à bas niveau (c'est à dire dans des petites fenêtres locales parcourant toute

l'image). Ces caractéristiques sont obtenues grâce aux descripteurs Speeded Up Robust Features (SURF) (Bay et al. (2008)) qui se basent sur la réponse d'ondelettes de Haar. 400 points sont extraits, décrits par un vecteur de dimension 128. Une représentation intermédiaire de l'image est obtenue en combinant ces caractéristiques locales. L'idée est d'obtenir une représentation compacte mais avec le minimum de pertes d'informations, c'est à dire une compression efficace. Trois approches sont expérimentées :

- L'approche des « Sacs de Mots Visuels » (*Bags Of Visual Words (BoVW)*). Cette méthode fonctionne en deux étapes. Chacun des descripteurs est d'abord traduit en un code puis les codes sont ensuite mis en commun pour former un unique vecteur représentatif. Un simple histogramme des descripteurs locaux répond par exemple à cette définition, mais des approches plus optimales existent.
- L'approche BossaNova, qui est une extension des BoVW. La première étape est inchangée, mais la mise en commun fait appel à l'histogramme des distances entre les descripteurs et le code associé. L'algorithme a pour sortie un histogramme par mot de code c_m . Pour chaque intervalle de cet histogramme, pour le code c_m considéré, on compte le nombre de descripteurs dont la distance à c_m se trouve entre certaines bornes fixées.
- L'approche par vecteurs de Fisher (Perronnin et al. (2010)), qui repose sur le principe des modèles par mélanges de Gaussiennes pour décrire l'ensemble de l'espace des descripteurs.

Pour chacune des trois représentations intermédiaires, un classifieur est utilisé pour discriminer les images pathologiques des non pathologiques. Dans leurs travaux, Pires et al. (2017) utilisent un SVM pour trouver l'hyperplan séparateur (la frontière de classification) dans chacun des trois espaces de représentations intermédiaires.

Cependant, cette approche se contente de faire du dépistage et n'indique pas l'avancée (le grade) de la maladie considérée. Dans le cadre de la gradation de la RD, un vaste concours a été lancé en 2015 sur la plateforme *Kaggle*¹. La tâche qui incombait aux participants était de classer plusieurs dizaines de milliers d'images selon l'avancée de la RD sur une échelle de 0 à 4 (absence de la maladie jusqu'à RD proliférative). Sans surprises, le podium de la compétition est occupé par trois approches reposant sur des réseaux de neurones convolutifs (bien qu'utilisant des architectures différentes). Face à cette hégémonie sans conteste, une étude scientifique a été publiée pour évaluer la capacité de déploiement de ces méthodes dans des applications de dépistage réelles (V. et al. (2016)) et concluait positivement.

Se basant sur l'analyse des résultats obtenus par les vainqueurs de la compétition, Roy et al. (2017) proposent une approche hybride, combinant une classification globale faite en com-

1. <https://www.kaggle.com/c/diabetic-retinopathy-detection>

binant des caractéristiques apprises automatiquement (comme précédemment) mais aussi une détection de lésions. L'architecture qui en découle est complexe : à partir d'une image d'entrée, deux branches divergent. La première branche consiste en un réseau de neurones classiques, prenant en entrée une image et produisant en sortie un vecteur de caractéristiques globales. La seconde branche découpe l'image en patchs. Pour chaque patch, un réseau de neurones convolutifs pré-entraîné sur une base de données de références (mais non liée au fond d'œil) est utilisé pour extraire un jeu de cartes de caractéristiques. Ce jeu est fourni à deux modules distincts : le premier est discriminatif, le second est génératif. Le modèle discriminatif utilise deux types d'algorithmes pour détecter la présence de lésions dans le patch, le premier étant un SVM et le second une forêt d'arbres décisionnels. La réponse de ces deux classifieurs forme l'histogramme discriminatif. Le second module fait appel aux vecteurs de Fisher pour décrire le jeu de caractéristiques sous la forme d'un mélange de Gaussiennes. Cela forme l'histogramme génératif. L'histogramme génératif, discriminatif et le vecteur de caractéristiques globales sont ensuite transmis à un dernier classifieur (une autre forêt d'arbres décisionnels) pour obtenir le grade de la maladie.

Cette approche, qui emprunte « au meilleur de chaque monde » (modèle discriminatif/génératif local et modèle global) obtient les meilleures performances.

2.4 Réseaux de neurones et apprentissage machine

Jusqu'à présent, l'accent a surtout été mis sur les méthodes dites « traditionnelles » (ou paramétriques) qui ne reposent pas sur des mécanismes d'apprentissage. En réalité, la reconnaissance de formes se base souvent sur ces mécanismes d'apprentissage. Likforman-Sulem and Barney-Smith (2013) propose de distinguer deux approches d'apprentissage : l'approche statistique et l'approche structurelle. La première consiste à rechercher les frontières des régions correspondantes aux différentes classes dans l'espace de représentation. La seconde repose sur l'apprentissage d'une grammaire à partir des structures des éléments constituant la base d'apprentissage, une grammaire représentant de façon générique une classe. Cette approche se prête bien à l'analyse par graphes.

Dans le domaine de l'image, l'approche la plus utilisée est souvent statistique, suivant des méthodes probabilistes (approche Bayésienne ou Markovienne, réseau de neurones, k plus proches voisins...) ou géométriques (analyse discriminante, arbre de décision, machine à vecteurs de supports...). Dans le premier cas, on modélise implicitement les frontières de décision en cherchant le modèle de distribution statistique des éléments de chaque classe et dans le second cas, on se concentre sur la recherche de l'hyperplan de séparation des classes.

Concernant l'approche statistique, la plupart des méthodes supervisées cherchent à modéli-

ser la distribution des classes y conditionnée aux caractéristiques \mathbf{x} observées sur un élément donné, notée $p(y|\mathbf{x})$ (Goodfellow et al. (2016)). Formellement, cela revient à maximiser l'espérance de la distribution de sortie du modèle $p_{\text{modèle}}(y|\mathbf{x})$ sur l'ensemble du jeu d'apprentissage \mathbf{X} associé aux labels \mathbf{Y} . Cela se fait en décrivant une fonction de coût sur cet ensemble d'apprentissage :

$$J(\omega) = \mathbb{E}_{\mathbf{X}, \mathbf{Y} \sim p_{\text{données}}} L(\mathbf{x}, y, \omega) \quad (2.9)$$

où ω correspond aux paramètres du modèle et L au coût par élément de la base d'apprentissage. De façon générique, l'apprentissage répond à l'objectif de minimisation de cette fonction de coût en optimisant les paramètres du modèle :

$$\omega^* = \underset{\omega \in \Omega}{\operatorname{argmin}} J(\omega) \quad (2.10)$$

En pratique, la recherche de l'optimal ω^* peut s'effectuer de plusieurs façons. Si la fonction J est convexe, une solution analytique peut s'obtenir (unique si J est strictement convexe). Ce n'est cependant généralement pas le cas, on recourt donc en général à d'autres méthodes. Par exemple, dans le cas d'un problème probabiliste, c'est à dire dans le cas où l'équation 2.9 revient à trouver les paramètres du maximum de vraisemblance du modèle (auquel cas, il s'agit d'un problème de maximisation et non de minimisation, mais la différence ne tient généralement qu'à un signe), l'algorithme d'espérance-maximisation décrit par Dempster et al. (1977) propose une approche itérative pour converger vers un maximum local.

Si la fonction J est dérivable, la technique de descente de gradients (introduite par Cauchy (1847)) propose de rechercher un optimum local, toutefois sans garantie qu'il soit global. Dans l'espace des paramètres de la fonction, on évalue le gradient de la fonction J au point courant ω_t . En progressant dans la direction opposée à ce gradient, on fait décroître la fonction et en appliquant itérativement le processus, l'algorithme permet d'atteindre un point stationnaire correspondant à un minimum. En revanche, si l'algorithme précise la direction à prendre, il ne permet pas d'obtenir la « distance » à parcourir dans cette direction. Dans son expression la plus simple, la descente de gradients utilise donc un pas d'apprentissage η (distance parcourue dans la direction du gradient) fixe :

$$\omega_{t+1} = \omega_t - \eta \nabla_{\omega} J(\omega_t) \quad (2.11)$$

La Figure 2.5 est extraite d'une simulation de descente de gradients dans un espace de dimensions deux sur une fonction arbitraire dont la dérivée est connue en tout point.² Dans

2. Le code source de la simulation, exécutable dans un environnement Jupyter, est disponible à l'adresse suivante : <https://github.com/ClementPla/IFT6135/blob/master/Convergence%20D%20surfaces.ipynb>

le cas de figure présenté, seul les pas d'apprentissages sont différents. Un pas d'apprentissage trop petit (Figure 2.5b) condamne le gradient à rester sur un point stationnaire correspondant à un minimum local.

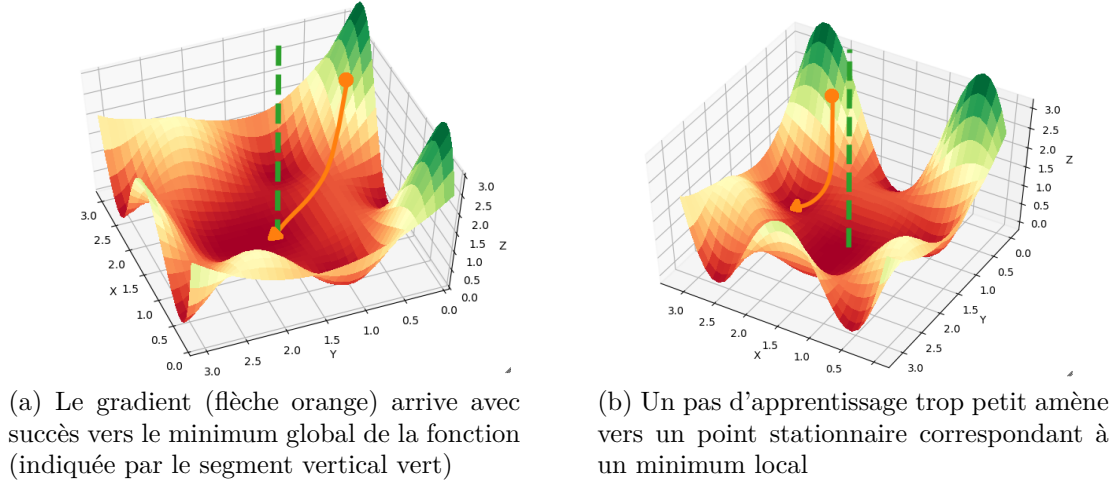


Figure 2.5 Descente de gradients : convergence vers un minimum global et vers un minimum local

De très nombreux travaux ont été publiés sur l'amélioration de la descente de gradients. Pour résister aux gradients bruités et offrir une plus grande robustesse dans des situations de surface à forte courbature, Polyak (1964) introduit un terme d'inertie (*momentum*) à l'algorithme. Pour cela, il introduit un vecteur de vitesse v , correspondant à la direction globale du gradient au fil des itérations. Ce vecteur n'est en fait que la moyenne glissante des gradients. La notion d'inertie permet de faire une analogie avec la mécanique Newtonienne : la descente de gradients est perçue comme la conversion d'énergie potentielle de pesanteur en énergie cinétique avec pertes dissipées dues aux frottements, dont le coefficient de friction vaudrait ϵ . Formellement, cela se traduit par l'algorithme suivant :

$$\mathbf{v}_{t+1} = \alpha \mathbf{v}_t - \epsilon \nabla_{\omega} J(\omega_t) \quad (2.12)$$

$$\omega_{t+1} = \omega_t + \mathbf{v}_{t+1} \quad (2.13)$$

L'équation 2.12 calcule l'inertie du modèle, modifiant l'équation 2.11 pour prendre en compte l'historique des précédentes orientations. Le scalaire ϵ peut s'interpréter de la même façon que le pas d'apprentissage η : il s'agit de la quantité de mouvement dans la direction du gradient. Le scalaire α , appelé coefficient d'inertie, contrôle le poids à accorder aux itérations passées. Sutskever et al. (2013) propose une variante de cet algorithme en consistant à évaluer le gradient de la fonction sous l'effet de la vitesse. Concrètement, cela revient simplement à

modifier l'équation 2.12 en :

$$\mathbf{v}_{t+1} = \alpha \mathbf{v}_t - \epsilon \nabla_{\omega} J(\omega_t + \alpha \mathbf{v}_t) \quad (2.14)$$

En pratique, cette approche est généralement plus stable et converge plus rapidement vers un minimum. En dehors des travaux de recherche sur l'inertie, de nombreuses publications traitent également de la recherche du pas d'apprentissage. Si, en théorie, il peut exister un pas optimal permettant d'atteindre le minimum en une itération, il requiert de calculer l'inverse de la matrice Hessienne, opération qui n'est pas réalisable en pratique. En apprentissage profond, des méthodes reposant sur un pas adaptatif sont par conséquent privilégiées ; l'intuition étant généralement de réduire le pas d'apprentissage dans les direction où le gradient est le plus fort et réciproquement. De cette façon, le gradient peut mieux coller à la surface de la fonction dans les cas où celle-ci est très accidentée, tout en étant capable de sortir des minima dans un plateau de très faible courbure (donc de faible gradient). L'une des approches les plus populaires est celle proposée par Kingma and Ba (2014) sous le nom de Adam. Elle intègre l'inertie (nommée moment du premier ordre) et corrige le pas d'apprentissage de façon inversement proportionnelle à la magnitude moyenne des précédents gradients (une moyenne glissante exponentielle étant utilisée). Kingma and Ba (2014) propose également de corriger le biais introduit par cette moyenne glissante, via un opérateur limitant la contribution des premières itérations. Le moment de premier ordre, non biaisé grâce à l'effet du dénominateur, est obtenu de façon similaire à l'équation 2.12 :

$$\mathbf{s}_{t+1} = \frac{b_1 \mathbf{s}_t + (1 - b_1) \nabla_{\omega} J(\omega_t)}{1 - b_1^t} \quad (2.15)$$

où le scalaire $b_1 < 1$ a une interprétation similaire à α de l'équation 2.12. Le dénominateur corrige le biais introduit par l'effet de moyenne glissante lors des premières itérations. Pour s'en convaincre, il suffit d'observer qu'il tend vers 1 quand t augmente, de sorte que rapidement, cette correction de biais n'a plus d'importance. Le moment de second ordre, lui, est obtenu par la moyenne quadratique glissante (et corrigée de son biais) :

$$\mathbf{r}_{t+1} = \frac{b_2 \mathbf{r}_t + (1 - b_2) (\nabla_{\omega} J(\omega_t) \odot \nabla_{\omega} J(\omega_t))}{1 - b_2^t} \quad (2.16)$$

où \odot représente le produit de Hadamard (multiplication de matrices élément par élément). \mathbf{r} représente donc la moyenne des carrés des gradients. L'équation de mise à jour des paramètres devient alors :

$$\omega_{t+1} = \omega_t - \eta \frac{\mathbf{s}_{t+1}}{\sqrt{\mathbf{r}_{t+1}} + \delta} \quad (2.17)$$

Le pas d'apprentissage η est donc bien corrigé par un facteur inversement proportionnel à l'amplitude moyenne ($\sqrt{\mathbf{r}}$) des gradients. δ est là pour assurer la stabilité numérique.

La Figure 2.6 illustre le résultat en simulation de l'algorithme Adam. L'effet de l'inertie est particulièrement visible, se traduisant par des oscillations en cercles concentriques de rayons décroissants autour du point minimum.

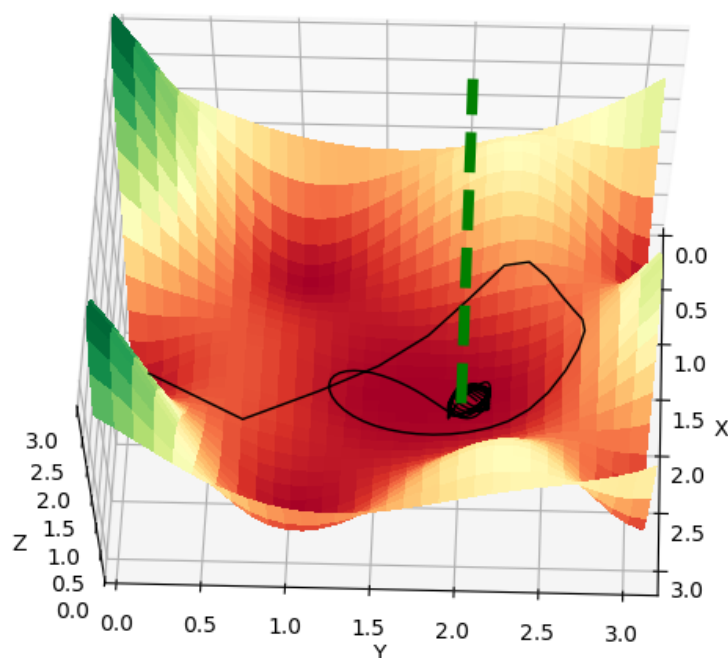
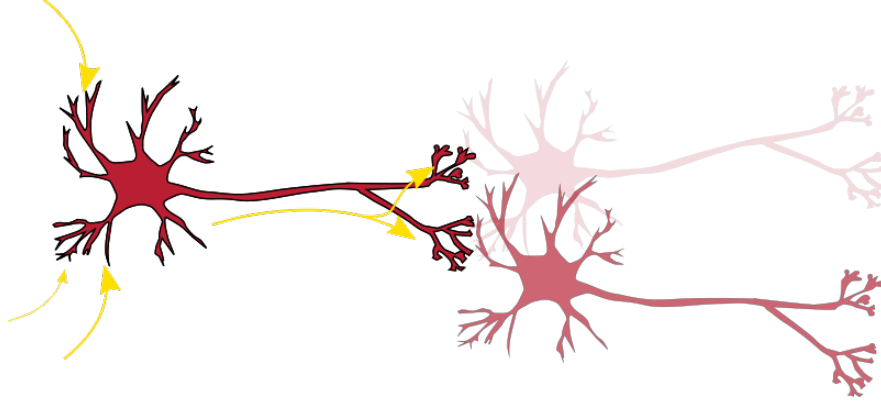


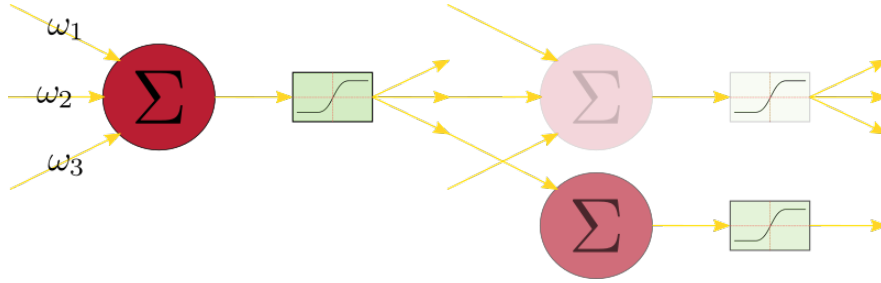
Figure 2.6 Trajet de la descente de gradient par la méthode Adam.

2.4.1 Le neurone formel

Le neurone artificiel tire son inspiration du modèle de fonctionnement des vrais neurones proposé par McCulloch and Pitts (1943). En résumé, un neurone possède des centaines de synapses, via lesquelles il est connecté aux autres neurones. Sa sortie dépend d'une certaine pondération de ses entrées, certains processus chimiques venant modifier le poids des interconnexions. Le neurone artificiel est conçu pour se comporter de manière similaire : sa sortie n'est que la somme de ses entrées. Pour coller encore davantage au modèle réel, dans lequel le signal ne se propage dans l'axone que s'il atteint une certaine valeur, la sortie du neurone artificiel est complétée par une fonction d'activation non linéaire visant à simuler cet effet seuil. La Figure 2.7 illustre le parallèle entre le neurone réel et l'artificiel.



(a) Le signal en provenance des autres neurones est transmis au corps cellulaire par les synapses via les dendrites. Le corps cellulaire transmet sa réponse par l'axone du neurone.



(b) La sortie du neurone artificiel est la somme pondérée des entrées passée à une fonction d'activation non linéaire.

Figure 2.7 Comparaison entre un réseau de neurone réel et son modèle artificiel

Mathématiquement, la sortie du neurone formel s'exprime comme étant :

$$y = f\left(\sum_i w_i \cdot x_i\right) \quad (2.18)$$

Un neurone ainsi formé fournit une frontière de séparation linéaire (si la fonction f est symétrique). Par exemple, en dimension 2, en supposant que la fonction d'activation soit symétrique, la frontière de séparation sera $x_2 = -(w_1/w_2)x_1$. Généralement trop limitée, on privilégie une frontière pouvant s'exprimer sous une forme affine. On ajoute pour cela un biais b à la somme des entrées du neurone. L'équation 2.18 devient alors :

$$y = f\left(\sum_i w_i \cdot x_i + b\right) \quad (2.19)$$

En dimension 2, la frontière de séparation devient alors effectivement une fonction affine : $x_2 = -(w_1/w_2)x_1 - b/w_2$. En pratique, le neurone formel n'a pas d'utilité à lui seul. En revanche, l'association de neurones formels, formant un réseau de neurones, généralement

composé de plusieurs couches placées en série, a fait le succès de l'apprentissage profond. Un réseau de neurones est caractérisé par son nombre de couches et le nombre de neurones par couche.

2.4.2 Les réseaux de neurones multicouches

Un réseau de neurones peut se voir comme une application $f_{\omega}(\mathbf{x}) = \mathbf{y}$, paramétrée par des poids ω , qui associe à un vecteur $\mathbf{x} \in \mathbb{E}$ un vecteur de correspondance $\mathbf{y} \in \mathbb{F}$, \mathbb{E} et \mathbb{F} étant des espaces vectoriels de dimensions différentes. Cette fonction est en réalité une composition d'une succession d'applications non-linéaires mises en cascade $f_{\omega}(\mathbf{x}) = g_{\omega_n} \circ g_{\omega_{n-1}} \circ \dots \circ g_{\omega_0}(\mathbf{x})$, chaque application g_{ω_j} correspondant à une couche dans le réseau indexée de 0 (l'entrée) à n (la sortie).

De fait, chacune de ces couches forme une représentation non linéaire de plus en plus complexe de l'entrée. Apparemment anodine, cette propriété est probablement une des clés du succès de l'apprentissage profond. À la différence des autres approches d'apprentissage machine entraînées à faire une classification à partir de caractéristiques extraites « à la main », les réseaux de neurones apprennent d'eux-mêmes à extraire les caractéristiques utiles à la discrimination entre classes. On parle alors d'apprentissage des représentations (Goodfellow et al. (2016)). La capacité à extraire des représentations non-linéaires est par ailleurs essentielle. En effet, sous cette condition de non-linéarité, les travaux de Cybenko (1989) ont démontré la capacité d'un réseau de neurones à une couche cachée à approximer n'importe quelle fonction continue dans un espace fermé dès lors que son nombre de neurones est assez grand. Ce résultat fondamental à la compréhension des réseaux de neurones est baptisé Théorème d'Approximation Universelle.

En traitement d'images, la brique essentielle des réseaux de neurones est la couche convolutive neuronale (Convolutional Neural Network (CNN)). Introduit par LeCun et al. (1990), les CNN répondent aux trois propriétés suivantes :

1. Une connectivité peu dense : Au lieu d'être connecté à l'ensemble de l'entrée qui le précède, chaque neurone n'est relié qu'à un petit sous-ensemble de celle-ci.
2. Le partage des poids : bien que tous les neurones ne soient pas tous connectés au même sous-ensemble de l'entrée, ils partagent tous les mêmes poids. Cela limite significativement la place en mémoire du modèle.
3. L'équivariance en translation : il s'agit d'une propriété intrinsèque de la convolution. L'équivariance d'une fonction f à une transformation T se traduit formellement de la façon suivante : $f(T(x)) = T(f(x))$

Ces trois propriétés sont satisfaites par l'utilisation de la convolution dont les paramètres du noyau K peuvent être vues comme équivalent aux poids du neurone formel. Pour une entrée E à Ω canaux on calcule la sortie S , possédant Γ canaux, de la façon suivante :

$$S(\gamma; i, j) = \sum_{\omega=1}^{\Omega} \sum_m^M \sum_n^N E(\omega; m, n) K(\gamma, \omega; i - m, j - n), \quad \gamma \in \{1, \dots, \Gamma\} \quad (2.20)$$

où $\mathbb{R}^{\Gamma \times \Omega \times M \times N}$ est le tenseur des poids de la couche. Γ représente la profondeur de la couche de sortie, Ω celle de la couche d'entrée et (M, N) représentent les dimensions spatiales du noyau. Pour satisfaire les deux premières propriétés des réseaux convolutifs, M et N sont généralement choisis petits (dans la majorité des architectures, $M = N \in \{1, 3, 5, 7\}$).

La convolution est une opération linéaire et n'est pas intéressante à entraîner en soi pour cette raison : en effet, la composition de plusieurs applications linéaires reste une application linéaire. Au mieux permettra-t-elle d'extraire une frontière de discrimination linéaire, ce qui est rarement suffisant. En conséquence, suivant le modèle du neurone formel, un biais et une fonction d'activation non-linéaire sont placés en sortie de la convolution :

$$\hat{S}(\gamma; i, j) = f(S(\gamma; i, j) + b_{\gamma}) \quad \gamma \in \{1, \dots, \Gamma\} \quad (2.21)$$

Les architectures de réseaux de neurones convolutives sont nombreuses. Les travaux de Le-Cun et al. (1998) ont posé les bases encore en vigueur dans la plupart des architectures récentes. L'architecture suit un modèle pyramidal, la profondeur des couches successives ne cessant de croître. Parallèlement, la résolution spatiale est régulièrement divisée par sous-échantillonnage. Le réseau apprend ainsi une représentation de plus en plus compacte de l'image d'entrée et devient invariant en translation. Cette représentation compacte est transmise à une succession de couches complètement connectées (par opposition aux couches convolutives qui ne sont que localement connectées) qui vont sortir un vecteur de classification. L'index de la valeur maximale de ce vecteur correspond à la classe prédite par le réseau. Cependant, cette approche ne permet pas de faire de la segmentation mais uniquement de la classification. Long et al. (2015) propose de remplacer les dernières couches connectées par le même type de structure pyramidale que précédemment mais inversée : les opérateurs de sous-échantillonnage sont remplacés par des opérations de sur-échantillonnage et la profondeur de chaque couche est progressivement réduite. La couche de sortie correspond à un tenseur de même dimension spatiale que l'image d'entrée et dont la profondeur est égale au nombre de classes. Ainsi, chaque canal de cette dernière couche représente la carte de segmentation de la classe correspondante. L'architecture permet ainsi de réaliser une segmentation sémantique de l'image d'entrée.

Plus récemment, He et al. (2016) ont proposé de connecter deux couches convolutives non-successives via des *connexions résiduelles* additionnant les sorties de ces couches. Ces connexions viennent en quelque sorte court-circuiter les couches placées en série. He et al. (2016) suggère de sauter au moins deux couches convolutives via ces connexions, ce qui se traduit, en reprenant les notations de l'équation 2.21, par :

$$Y_{CR} = \hat{S}_n(\gamma; i, j) + \hat{S}_{n-2}(\gamma; i, j) \quad (2.22)$$

où Y_{CR} est la sortie de la connexion résiduelle et où l'indice des sorties de convolution correspond à la position de la couche dans le réseau (\hat{S}_0 étant l'image d'entrée). Les connexions résiduelles apportent deux avantages :

- Elles facilitent l'apprentissage de la fonction identité
- En conséquence, elles facilitent également la propagation du gradient de la sortie vers l'entrée du réseau.

Grâce à des deux propriétés, il devient possible d'entraîner des réseaux beaucoup plus complexes (c'est à dire composés de nombreuses couches cachées) et donc gagner en capacité de généralisation.

Dans le domaine biomédical, l'architecture U-Net présentée par Ronneberger et al. (2015) a rencontré un vif succès. En insérant des « connexions passerelles » (*skipped connections*) entre les couches convolutives de même résolution, cette architecture a réussi à segmenter des structures neuronales sur des images prises au microscope électronique alors même que la base d'apprentissage ne comportait que quelques dizaines d'images au plus. L'article joint au mémoire s'appuie sur une structure de réseau similaire au U-Net. Cette capacité à généraliser sur aussi peu de données (justifiée par un gradient se propageant plus facilement grâce aux connexions passerelles) est naturellement d'un grand attrait pour la communauté biomédicale devant faire face à une certaine rareté des données annotées.

Pour illustrer cette rareté, la section 2.5 recense toutes les bases de données publiques existantes liées à des applications reliées au fond d'œil.

2.4.3 Métriques d'évaluation de performance

La question de l'évaluation de la performance des algorithmes de segmentation (et plus généralement de classification) n'est pas anodine. En effet, il n'existe pas de métrique universelle communément admise à laquelle se référer. Dans le domaine médical, l'évaluation de la qualité de la segmentation se fait toujours par rapport à une référence, celle-ci étant réalisée par le clinicien. Cependant, il n'est pas exclu que le clinicien fasse des erreurs. A ce titre, les bases de données les plus rigoureuses fournissent des références produites par au moins

deux experts. Un mécanisme de vote ou de consensus peut dès lors être mis en place pour construire la « vérité terrain ». Pour autant, cela ne garantit pas qu'elle soit idéale. Dans le cas des lésions, ce cas de figure arrive typiquement si tous les experts se focalisent davantage sur la localisation que sur le marquage précis des frontières. Ainsi, pour pouvoir comparer la sortie d'un algorithme avec la vérité terrain, une bonne métrique se doit de refléter les sources d'erreurs suivantes :

1. La sur-segmentation : l'algorithme identifie des segments systématiquement plus grands que ceux de la référence.
2. La sous-segmentation : inversement, l'algorithme identifie des segments systématiquement plus petits que ceux de la référence.
3. L'imprécision sur les frontières, en partie due à l'appréciation subjective des experts humains.
4. Un écart sur le nombre de régions segmentées.

Dans sa revue de littérature, Zhang (1996) propose trois catégories de méthodes pour l'évaluation des performances de segmentation : les méthodes analytiques (*Analytical Methods*), les méthodes basées sur une évaluation empirique de la justesse (*Empirical Goodness Methods*) et les méthodes basées sur l'analyse du désaccord empirique (*Empirical Discrepancy Methods*). Les deux premières ne requièrent pas de références (évaluation non supervisée). La méthode analytique évalue un algorithme non pas en fonction de sa sortie mais plutôt en considérant les principes utilisés, la complexité de mise en place, son erreur théorique, etc. L'évaluation empirique de la justesse ne nécessite pas non plus de références : elle évalue la performance de l'algorithme en quantifiant la similarité entre régions segmentées. Zhang (1996) distingue trois propriétés exploitables pour évaluer la justesse empirique de la segmentation :

- L'uniformité intra-région.
- Le contraste inter-régions.
- La forme de la région.

En pratique, ces approches non-supervisées sont peu utilisées. La première n'a d'intérêt pratique que dans des cas très spécifiques se rapprochant davantage de problèmes d'ingénierie que scientifiques. La seconde est critiquable du fait que le critère d'évaluation risque d'être le même que le critère de segmentation. Une justesse mesurée en fonction du contraste inter-régions sur un algorithme de segmentation se basant justement sur ce critère n'est guère indicatrice des performances de l'algorithme.

Une approche d'évaluation supervisée, donc basée sur une référence, garantit une évaluation non biaisée. Là encore, plusieurs approches existent. Entre autres, la mesure peut se faire au niveau du pixel, de la position des pixels mal-classifiés ou encore sur la différence du nombre

de régions segmentées. Ces deux dernières approches rentrent surtout en compte dans la mesure de performance de détection. Pour le fond d'œil, elles sont souvent employées pour évaluer les performances d'identification des toutes petites lésions (micro-anévrismes) : en effet, l'accord sur les frontières importe peu dans ce cas, à la différence de l'accord sur la position et la présence des lésions. Pour l'accord au niveau du pixel, plusieurs métriques existent pour mesurer l'accord (ou le désaccord) entre les deux observateurs. Pour les introduire, on pose S (pour *sortie*) l'image segmentée par l'observateur 1 (l'algorithme), V (pour *vérité*) la référence apportée par l'observateur 2 (le médecin). Enfin, notons S_1 (resp. V_1) les régions de S (resp. V) notées comme faisant partie de l'avant-plan et S_0 (resp. V_0) celles étant de l'arrière plan. Ceci permet de définir des sous-ensembles de l'image segmentée :

- VP (vrai positif) = $S_1 \cap V_1$
- FP (faux positif) = $S_1 \cap V_0$
- VN (vrai négatif) = $S_0 \cap V_0$
- FN (faux négatif) = $S_0 \cap V_1$

De là, les statistiques de matrice de contingence permettent d'obtenir, pour une image entière, les mesures de performance usuelles :

1. La **sensibilité** : $Se = \frac{|VP|}{|V_1|} = \frac{|VP|}{|VP|+|FN|}$
2. La **spécificité** : $Sp = \frac{|VN|}{|V_0|} = \frac{|VN|}{|VN|+|FP|}$
3. La **précision** : $Pr = \frac{|VP|}{|S_1|} = \frac{|VP|}{|VP|+|FP|}$
4. L'**exactitude** (*accuracy*) : $Ex = \frac{|VP|+|VN|}{|VP|+|VN|+|FP|+|FN|}$

Où l'opérateur $|x|$ est le cardinal de l'ensemble x . Pour la segmentation d'images, notamment en présence d'un grand déséquilibre de classes (en général, la classe 0 -le fond- est bien plus représentée que la classe 1), on s'intéresse généralement au doublet précision-sensibilité. Une grande précision s'opère généralement (peu de faux positifs) au détriment d'une bonne sensibilité (pas assez de vrai positifs). Il est possible de combiner ces deux métriques pour obtenir la F-mesure, introduite par Van Rijsbergen and Blair (1979) :

$$F_\beta = \frac{(\beta^2 + 1)Pr \cdot Se}{\beta^2 \cdot Pr + Se} \quad (2.23)$$

Le paramètre β contrôle l'importance accordée à la sensibilité par rapport à la précision ($\beta > 1$ privilégie la sensibilité). De manière générale, un système de classification met toujours en jeu un compromis entre la quantité de fausses détections et la quantité de détections manquées. La plupart des algorithmes dépendent d'un jeu de paramètres influant la première au détriment de la seconde et réciproquement. En discrétisant l'espace de ces paramètres, Spackman (1989) a montré qu'il était possible de tracer la courbe d'efficacité du récepteur

(dit Receiver Operating Characteristic (ROC)). Celle-ci place en ordonnées le taux de vrais positifs (sensibilité) et en abscisses le taux de faux positifs (1-spécificité) (voir la Figure 4.6 pour exemple). L'aire mesurée sous la courbe est une indication de la performance du système pour l'ensemble des paramètres considérés. C'est donc une mesure permettant de comparer facilement deux systèmes différents.

En traitement d'image, on privilégie des métriques géométriques à interprétation visuelle. De fait, une métrique très souvent utilisée est la distance de Jaccard, introduite par Jaccard (1901). Elle est définie comme :

$$D_{Jaccard} = 1 - \frac{|S_1 \cap V_1|}{|S_1 \cup V_1|} \quad (2.24)$$

Cette distance peut se réinterpréter dans le cas binaire de la façon suivante³ :

$$D_{Jaccard} = \frac{|S_1 \cup V_1| - |S_1 \cap V_1|}{|S_1 \cup V_1|} = \frac{|VP| + |FP| + |FN| - |VP|}{|VP| + |FP| + |FN|} \quad (2.25)$$

$$= \frac{|FP| + |FN|}{|VP| + |FP| + |FN|} = \frac{|FP| + |FN|}{|V_1| + |FP|} \quad (2.26)$$

Il s'agit donc du rapport entre le nombre de pixels incorrectement classifiés et le nombre de pixels positifs et/ou segmentés comme tels.

Dans le domaine de la segmentation médicale, une métrique très utilisée est l'index de Sørensen–Dice (Dice (1945)), définie par :

$$Dice = \frac{2|S_1 \cap V_1|}{|S_1| + |V_1|} = \frac{2|VP|}{2|VP| + |FP| + |FN|} \quad (2.27)$$

Le coefficient de Sørensen–Dice est lié à la distance de Jaccard par la relation suivante :

$$D_{Jaccard} = 1 - \frac{Dice}{2 - Dice} \quad (2.28)$$

Par ailleurs, il est égal à la F-mesure pour $\beta = 1$. La distance de Jaccard ou l'indice de Sørensen–Dice sont tous les deux basés sur des considérations ensemblistes et géométriques. De ce point de vue là, des méthodes plus élaborées existent. Huang and Dom (1995) propose un variante directionnelle de la distance de Hamming (originellement conçue pour mesurer la similarité entre deux chaînes de caractères). Cardoso and Corte-Real (2005) propose une mesure de distance de partition définie comme le nombre minimum d'éléments de l'image qui devraient être supprimés pour que S et V (les deux partitions) soient identiques.

3. On peut noter que la librairie Python Scikit-Learn, très répandue, contient une erreur d'implémentation et confond l'exactitude avec l'indice de Jaccard

En plus des mesures de similarité statistique et géométrique, la segmentation peut également s'évaluer via des considérations probabilistes. Les deux cartes S et V sont de ce fait considérées comme des distributions dont on mesure la dissimilarité, notion qui ne prend sens qu'en définissant des métriques sur des distributions. Il en existe plusieurs : l'une des plus utilisées est la divergence de Kullback-Leibler (Kullback and Leibler (1951)), appartenement initialement au domaine de la théorie de l'information mais souvent utilisée dans les modèles d'apprentissage profond sous la forme du maximum de vraisemblance, ou encore la métrique de Wassertein (Olkin and Pukelsheim (1982)). Celle-ci peut se décrire intuitivement comme étant le coût (produit de la quantité par la distance) nécessaire pour transformer une pile de cailloux en une autre pile donnée, les piles représentant les deux distributions.

Dans l'article joint à ce mémoire, le coefficient Kappa de Cohen κ (Cohen (1960)) est utilisé comme objectif d'entraînement. Celui-ci mesure l'accord relatif entre deux observateurs (ici, S et V) en prenant en compte que l'accord peut être dû au hasard. Pour illustrer cette notion d'accord dû au hasard, imaginons que deux observateurs tirent 100 fois une pièce de monnaie très déséquilibrée (par exemple, elle tombe sur face 90% du temps). Si on souhaite mesurer l'accord entre les deux séquences observées, l'exactitude (par exemple) sera une mauvaise métrique : si, connaissant le biais de la pièce, un observateur fausse l'expérience et ne décide de noter que des faces, cette métrique indiquera quand même un score de 90% sans que cela ne traduise une quelconque corrélation avec l'autre observateur. Le κ propose donc de corriger ce biais. Ainsi, un accord reposant sur le hasard obtiendra un $\kappa = 0$, un accord parfait un $\kappa = 1$ et un désaccord sur chaque issue de l'expérience (c'est à dire sur chaque pixel dans le cas de la segmentation, $S_0 = V_1$) obtiendra un $\kappa = -1$.

Formellement, le κ se définit comme :

$$\kappa = \frac{P_{observateurs} - P_{hasard}}{1 - P_{hasard}} \quad (2.29)$$

où $P_{observateurs}$ est l'accord relatif entre les deux observateurs et P_{hasard} est la probabilité d'un accord aléatoire. Dans le cas binaire, on a :

$$P_{observateurs} = \frac{|VP| + |VN|}{|VP| + |VN| + |FP| + |FN|} \quad (= Ex) \quad (2.30)$$

En terme de géométrie, cela se traduit par :

$$P_{observateurs} = \frac{|S_1 \cap V_1| + |S_0 \cap V_0|}{|V|} \quad (2.31)$$

La probabilité d'accord aléatoire est donnée par :

$$P_{hasard} = \frac{|VN| + |FP|}{|VP| + |VN| + |FP| + |FN|} \cdot \frac{|VN| + |FN|}{|VP| + |VN| + |FP| + |FN|} + \quad (2.32)$$

$$\frac{|VP| + |FP|}{|VP| + |VN| + |FP| + |FN|} \cdot \frac{|VP| + |FN|}{|VP| + |VN| + |FP| + |FN|} \quad (2.33)$$

Ce que nous pouvons ré-écrire sous une forme condensée :

$$P_{hasard} = \frac{|S_0|}{|S|} \cdot \frac{|V_0|}{|V|} + \frac{|S_1|}{|S|} \cdot \frac{|V_1|}{|V|} = \frac{|S_1| \cdot |V_1| + |S_0| \cdot |V_0|}{|V|^2} \quad (2.34)$$

car $|V| = |S|$. En combinant les équations 2.29, 2.31 et 2.34, on obtient :

$$\kappa = \frac{|V| \cdot (|S_1 \cap V_1| + |S_0 \cap V_0|) - |S_1| \cdot |V_1| - |S_0| \cdot |V_0|}{|V|^2 - |S_1| \cdot |V_1| - |S_0| \cdot |V_0|} \quad (2.35)$$

Le κ est un indicateur intéressant car robuste au déséquilibre de classes. On peut démontrer ce fait en considérant les deux cas suivant : un équilibre parfait et un déséquilibre profond. Dans le premier cas, on a $|V_0| = |V_1| = |V|/2$. En remarquant que $|S_1| + |S_0| = |S| = |V|$, on obtient :

$$\kappa_{equilibre} = \frac{2 \cdot (|S_1 \cap V_1| + |S_0 \cap V_0|) - |V|}{|V|} = 2 \cdot P_{observateurs} - 1 \quad (2.36)$$

Dans le cas d'un équilibre parfait, la probabilité d'être avantagé lors d'un tirage au hasard n'existe pas ($P_{hasard} = 0.5$). Ainsi, la performance se résume logiquement à l'accord relatif entre observateurs.

Dans le cas d'un très fort déséquilibre de classes, avec $|V_0| \gg |V_1|$, on peut émettre l'hypothèse que $|V_0| \approx |V|$ et $\frac{|V_1|}{|V_0|} \approx 0$. On obtient donc en première approximation de l'égalité 2.35 :

$$\kappa_{desequilibre} \approx \frac{(|S_1 \cap V_1| + |S_0 \cap V_0|) - |S_0|}{|V| - |S_0|} = \frac{|S_1 \cap V_1|}{|S_1|} - \frac{|S_0| - |S_0 \cap V_0|}{|S_1|} \quad (2.37)$$

$$= \frac{|VP|}{|VP| + |FP|} - \frac{|FP|}{|VP| + |FP|} = 2 \cdot \frac{|VP|}{|VP| + |FP|} - 1 \quad (2.38)$$

Le premier terme de l'équation 2.38 correspond au double de la précision. On vérifie bien que pour un nombre de vrai positifs égal à celui de faux positifs, le κ est nul et qu'il est inférieur à 0 si $|VP| < |FP|$. On constate également que la spécificité n'est plus prise en compte. Celle-ci n'est en effet pas indicatrice des performances de l'algorithme : si celui-ci prédisait tout en négatif, le taux de pixels correctement classifiés serait très élevé, mais pour autant la classification ne serait pas meilleure qu'un tirage au sort suivant la distribution des deux classes. On perçoit donc intuitivement l'inutilité de prendre en compte la spécificité.

La démonstration peut donc se résumer de la sorte : dans un cas équilibré, le κ se ramène à l’exactitude. Dans le cas déséquilibré, il se ramène à la précision.

Par ailleurs, si on fait l’hypothèse $|FP| \neq 0$, on peut réécrire l’équation 2.38 sous la forme suivante :

$$\kappa_{desequilibre} = \frac{2 \cdot |VP|/|FP|}{|VP|/|FP| + 1} - 1 \quad (2.39)$$

Ce qui permet de tracer l’évolution du $\kappa_{desequilibre}$ en fonction du rapport entre le nombre de vrais positifs et faux positifs (Figure 2.8).

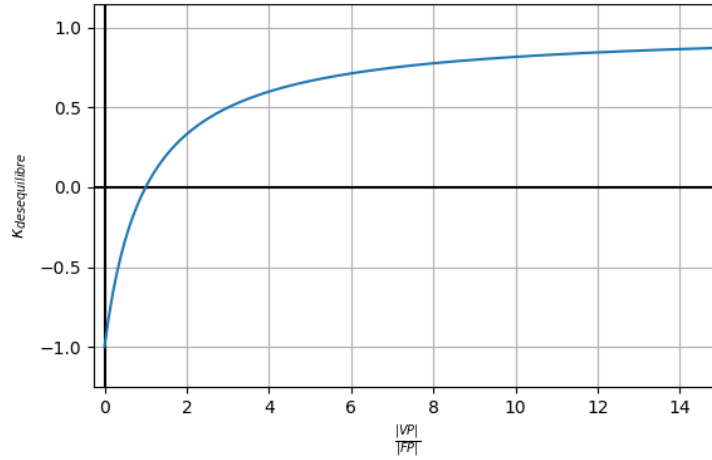


Figure 2.8 Évolution du κ dans un contexte de fort déséquilibre de classes en fonction du rapport VP/FP

2.5 Bases de données publiques

Les données constituent l’enjeu essentiel des approches reposant sur l’apprentissage machine. Seule une quantité importante de donnée permet d’assurer la capacité de généralisation des algorithmes développées. Au cours des deux dernières décennies, plusieurs bases de données ont été rendues publiques. Ces bases de données diffèrent entre elles par le type d’annotations qui accompagnent les images. Par ailleurs, les images elles-même ne sont pas toutes acquises dans les mêmes conditions à la même résolution. Le tableau 2.1 propose une liste exhaustive des bases de données publiques existantes ainsi que le type d’annotations (vérité terrain) qu’elles apportent. Bien qu’ils en existent un nombre relativement conséquent, très peu d’entre elles proposent une annotation complète des lésions. Pour celles qui le font, elles se concentrent généralement sur un type bien spécifique de lésions ; aucune ne propose une vérité terrain complètement exhaustive.

Tableau 2.1 Comparatif des bases de données publiques de fond d’œil

Nom	Référence	Nombre d’images	Type de vérité terrain.
STARE	Hoover et al. (2000)	402	Diagnostic de maladie + divers masques sur quelques images
DRIVE	Staal et al. (2004)	40	Masque sémantique du réseau vasculaire
ONHSD	Lowell et al. (2004)	99	Masque du disque optique
DIARETDB0	Kauppi et al. (2007)	130	Texte indiquant la présence de lésions par image
DIARETDB1		89	Annotations des régions de l’image contenant des lésions
DRIONS-DB	Carmona et al. (2008)	110	Masque du disque optique
ROC	Niemeijer et al. (2010)	100	Coordonnées des micro-anévrismes
INSPIRE-stereo	Tang et al. (2011)	30 (stéréo)	Profondeur obtenue par OCT
INSPIRE-AVR	Niemeijer et al. (2011)	40	Masque du réseau vasculaire et AVR
RIM-ONE	Fumero et al. (2011)	159	Masque du disque optique
E-OPHTA (EX)	Decencière et al. (2013)	82	Masque des exsudats durs
E-OPHTA (MA)		381	Masque des micro-anévrismes
HRF - segmentation	Köhler et al. (2013)	45	Masque du réseau vasculaire
HRF - quality		18 (mauvaise qualité)	18 images reprises (bonne qualité)
MESSIDOR-2	Decencière et al. (2014)	874×2 (paire D/G)	Grade de la RD et risque de OMD.
Drishti-GS1	Sivaswamy et al. (2015)	101	Masque du disque optique et de l’excavation papillaire
EyePACS - Kaggle	EyePACS (2015)	88702	Grade de la RD.

CHAPITRE 3 PROBLÉMATIQUE ET OBJECTIFS SPÉCIFIQUES

La revue de littérature réalisée permet de synthétiser l'état de la recherche et des enjeux actuels et a servi à orienter nos travaux de recherches menés. Pour résumer, l'imagerie de fond d'œil permet de déceler les symptômes de plusieurs maladies de façon anticipée et de classer la maladie selon son avancée. Or, si l'acquisition est aisée, l'analyse est une opération longue et fastidieuse pour l'ophtalmologiste, ce qui freine l'accès aux soins à un large de pan de la population là où le dépistage devrait être systématique. C'est pourquoi la recherche en vision par ordinateur appliquée aux images de la rétine est extrêmement active. Parallèlement, le secteur de l'intelligence artificielle a connu depuis quelques années des avancées majeures grâce à l'émergence de l'apprentissage profond et l'utilisation de réseaux neuronaux résolvant des tâches de plus en plus complexes.

Malgré cela, des verrous techniques subsistent encore et empêchent un réel déploiement clinique des solutions algorithmiques :

- Les systèmes de dépistage de la maladie souffrent d'un des deux défauts suivants : ou bien ils ne sont pas assez polyvalents et ne permettent pas de détecter toutes les lésions nécessaires au diagnostic ou bien ils manquent de transparence, en établissant un diagnostic fondé sur des critères non-interprétables.
- La quantité de données, encore trop limitée, freine le déploiement de solutions basées sur l'apprentissage profond. Or, cette technique a fait ses preuves par ailleurs en démontrant sa capacité de généralisation à de nombreux cas.

A la vue de ces contraintes, le défi est de réussir à collecter des données en quantité suffisante et riches en informations cliniques pour entraîner des réseaux de neurones capables de réaliser la segmentation de toutes les lésions. Pour accélérer le processus d'annotations, nous proposons de concevoir un système de pré-annotations qui pourra se perfectionner au fur et à mesure que la base de données d'apprentissage s'étoffera. Plus qu'une simple aide à l'annotation, ce système représente en lui-même un projet de recherches important, car une approche simultanée de la segmentation de plusieurs structures de la rétine n'a jamais été explorée.

Ce système, bien que voué à s'améliorer, se doit d'arriver à des résultats de segmentation de lésions convaincants dès son intégration à la plateforme d'annotations afin d'accélérer la démarche du clinicien (limitant ainsi son travail, dans le meilleur des cas, à une simple validation de la pré-annotation). Malgré tout, un tel système reste difficile à entraîner en l'état des bases de données publiques. Nous proposons d'explorer des pistes d'entraînements capables d'exploiter des bases de données non annotées. La principale problématique de notre travail

de recherche peut donc se résumer de la façon suivante :

Question de recherche : Peut-on exploiter les données de gradations de la maladie au niveau de l'image pour améliorer les performances de segmentation de lésions au niveau du pixel ? L'hypothèse sous-jacente à cette problématique est qu'il existe une corrélation forte entre le grade de la maladie et la présence de lésions dans l'image. Cette corrélation est intuitive : le diagnostic du grade de la maladie est fondé sur le repérage des lésions par le clinicien. Par conséquent, un réseau de neurones capable de diagnostiquer le grade de la maladie est donc théoriquement capable d'identifier ces lésions. Corollairement, si un système dédié à la segmentation est capable de tirer parti de la combinaison des informations au niveau de l'image entière et au niveau du pixel, il devrait offrir des meilleures performances, ayant été entraîné sur des bases de données bien plus larges.

Pour vérifier cette hypothèse, deux objectifs spécifiques ont été fixés :

1. Développer une architecture de réseaux de neurones cherchant toutes les lésions. Deux recherches parallèles ont été expérimentées : la segmentation multitâche d'une part et l'apprentissage faiblement supervisé. Ce dernier vise à tirer parti des grandes bases de données de gradations (la vérité terrain ne consiste qu'en un scalaire indiquant le degré de la maladie) pour améliorer la segmentation (qui requiert normalement une vérité terrain par pixels).
2. Proposer une plateforme d'annotations pour enrichir les bases de données d'apprentissage.

CHAPITRE 4 ARTICLE 1: A NOVEL WEAKLY SUPERVISED MULTITASK ARCHITECTURE FOR RETINAL LESIONS SEGMENTATION ON FUNDUS IMAGES

Clément Ployout, Renaud Duval and Farida Cheriet.

Submitted to IEEE- Transactions on Medical Imaging

4.1 Introduction

With recent breakthroughs in machine learning, the promise of automatic disease screening is progressively getting more tangible. In the field of ophthalmology, the development of deep learning, as well as the ease of acquisition of retinal fundus images, has already led to the approval of an automatic diabetic retinopathy (DR) grading system by US health authorities FDA (2018). Fundus images give access to a highly detailed 2D representation of the surface of the retina, generally centered on the macula or the optic disc and containing several diagnostically relevant biomarkers. An abnormality in the retina can either be a manifestation of an eye disease (e.g. age-related macular degeneration, DR), a systemic disease (e.g. hypertensive retinopathy) or even a trauma (e.g. traumatic macular hole). From a clinical point of view, the presence or absence of a given lesion, as well as its shape, texture and location are essential indicators for grading the progression and the severity of diseases. For example, early detection of microaneurysms (μA) is essential as their presence in the early stages of DR is the criterion to distinguish between non-healthy and healthy fundi. Therefore, the detection of μA has been the object of an online challenge Niemeijer et al. (2010). As μA are usually only a few pixels wide, the submitted approaches focus on detection rather than segmentation and share the same steps: preprocessing, candidates extraction, feature selection and candidates classification. Due to the circular shape of the μA , candidates extraction can be done using morphological operations Niemeijer et al. (2005); Cree et al. (1997), followed by match-filtering with a prior on the Gaussian profile of μA as introduced in Spencer et al. (1992). However, this method remains sensitive to background noise, so an approach based on local contrast independent of the size and shape of the regional minimum was proposed in Seoud et al. (2016). Features are then extracted for each candidate (morphological and intensity based in Niemeijer et al. (2005); Cree et al. (1997); Spencer et al. (1992), dynamic in Seoud et al. (2016)) and classified (Random Forest in Seoud et al. (2016), KNN in Niemeijer et al. (2005)). In addition to match-filtering, template-matching approaches with generalized Gaussian functions were also proposed in Quéllec et al. (2008). In order to counter-balance image illumination variations and noise, the authors propose to work in the wavelet domain.

Nonetheless, all those approaches rely on hyper-parameters that need to be fine-tuned, which might compromise the generalization capacity of the system. Therefore, fully trainable approaches have also been experimented Haloi (2015); Budak et al. (2017). In Haloi (2015), pixel-wise classification is done using a CNN. The network is trained on image patches and results are shown to improve with the use of non-uniform sampling and foveation. In Budak et al. (2017), candidates are extracted by analyzing the intensity profile along a spiral reading of a slicing window. A convolutional network is then used for automatic feature extraction and classification of each candidate patch, trained with reinforcement sample learning.

Concerning bright lesion segmentation, the literature mostly focuses on detection and/or segmentation of exudates. As for μA , a common approach is to identify candidates and refine the results through further analysis. For example, in Walter et al. (2002); Sreng et al. (2013); Zhang et al. (2014), candidates extraction is achieved with morphological operations and the refinement by applying maximum entropy thresholding Sreng et al. (2013) or Random Forest classification Zhang et al. (2014). In Liu et al. (2017), an alternative approach is proposed in which a Random Forest classifies each patch depending of the presence or absence of lesions. Positive patches are then forwarded to a segmentation module that refines the boundaries based on local variance, size and local contrast. The reviews in Kaur and Mittal (2018); Joshi and Karule (2018) give exhaustive overviews of the state of the art for exudate segmentation and in addition Kaur and Mittal (2018) proposes a generalized method to improve on the previous works. Their approach relies on adaptive k-means clustering followed by dynamic decision thresholding. Nonetheless, the proposed method necessitates optic disc segmentation and vessels removal. CNN-based approaches have also been experimented, as in Chudzik et al. (2018). Interestingly, to be able to successfully train their network, the authors pre-trained it on a different but related task, namely the segmentation of μA . Transfer-learning and finetuning were then applied to exudates segmentation. The work in Niemeijer et al. (2007) proposes an approach based on two steps: the first step identifies potential bright lesions, which are then classified according to their type (drusen, exudates, or cotton-wool spots) in a second step. This was the first work on joint segmentation of all types of bright lesions. The first classifier used is a k-NN while the second step uses a linear discriminant classifier. This work was further extended in van Grinsven et al. (2016) to be integrated in a screening program able to identify the presence or absence of drusen and exudates in retinal images, as both are related to different diagnoses.

In addition to pixel-wise identification of lesions, more global approaches have also been proposed. The work in V. et al. (2016) has demonstrated the ability of deep learning to do screening at a large scale for diabetic retinopathy and macular edema. However, an important limitation of their approach is that the deep network acts like a black box, meaning that

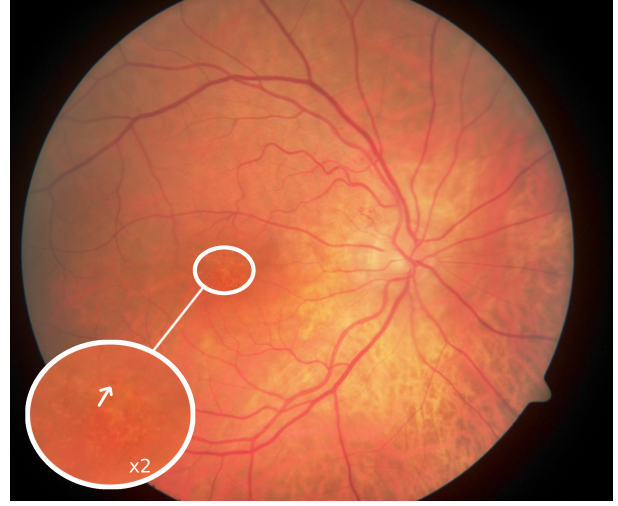
it discriminates over features that are not well understood. In Roy et al. (2017), the authors show that including discriminative and generative pathology histograms and combining them with CNN features significantly improves the performance of their grading system. In Quellec et al. (2017), a novel and promising approach is explored: starting from a convolutional network trained on a screening task (referable VS non referable), it is shown that by propagating the derivative of the prediction back to the input, the heat map obtained revealed lesions with a similar accuracy than a pixel-wise trained convolutional network. Nonetheless, the network cannot be used as it, because simply taking the derivative lead to a high number of false positives pixels caused by drafts. Therefore, the authors propose a novel finetuning procedure based on a three steps (back, forward and back again) propagation that limits the number of false positives.

A major limitation of global grading systems is that they are usually limited to a single diagnosis (for example, DR screening). As such, it is difficult to anticipate how the system will react when presented with pathological images affected by other diseases. In the worst case scenario, it might classify a pathological image as healthy, which could have dangerous consequences for the patient. Figure 4.1 presents two cases of non-diabetic patients who exhibit retinal lesions that could be misclassified as DR-related.

In light of this problem, the present work focuses on the segmentation of retinal lesions independently from a particular diagnostic task, using a neural network approach. We divide the biomarkers into two classes, *red lesions* (including haemorrhages of all shapes and microaneurysms) and *bright lesions* (hard and soft exudates, drusen, cotton-wool spots), that are general enough that our system behaves as a pixel-wise abnormality detector. From a generalization point of view, this allows the architecture to be sensitive to rare lesions, without compromising on the quality of a posterior grading system, as those markers remain diagnostically relevant. In our previous work Playout et al. (2018), we have shown that multi-tasking and hard parameter sharing are suitable for retinal lesion segmentation. Intuitively, this can be explained by the fact that bright and red lesions, while being very different in shape, texture and size, are not uncorrelated: the presence of one type of lesion is a strong sign for the potential presence of the other type. The main contribution of the present work is to propose a novel procedure for generalizing the network training by using labels at the image level, and we show that this significantly improves the system’s performance in term of precision and sensitivity on different databases. The paper is organized as follows. In section 4.2, we provide a detailed description of the methodology and the training process. Section 4.3 describes the experimental validation setup and Section 4.3.3 reports the results which are then discussed in Section 4.4.



(a) Two cotton-wool spots and a retinal hemorrhage.



(b) A microaneurysm near the fovea and small vascular anastomosis near the optic nerve.

Figure 4.1 Examples of non diabetic patients exhibiting retinal lesions. In figure 4.1a, the patient exhibits a secondary retinal hemorrhage consecutive to a pancytopenia caused by a myelodysplastic syndrom. In figure 4.1b, the patient recovers from BRVO (Branch Retinal Vein Occlusion) and presents a small microaneurysm near the fovea.

4.2 Proposed method

Our methodology relies on a fully convolutional architecture trained following a novel procedure that combines training at a pixel level and weakly supervised training with label at an image level. The proposed system takes as an input an RGB image and outputs two probability maps for red and bright lesions segmentation. First, it determines the circular region of interest in order to apply spatial calibration. Then, a preprocessing step is applied to remove background noise, correct the non-uniformity of the illumination and enhance the contrast. Third, our multi-task architecture is trained on a supervised way. In a fourth step, the training is improved and generalized using weak supervision. We detail each step in the following section.

4.2.1 Preprocessing

Fundus images can present uneven illumination, high variability in terms of contrast and may be subject to important background noise. We address those issues by applying successive steps inspired by Seoud et al. (2016). Nonetheless, our method differs in that it preserves the color consistency of the input image.

4.2.1.1 ROI estimation

The ROI is estimated by thresholding the median filtered red channel of the fundus. The median filter is computed with a large window, therefore preventing the thresholding from eliminating any darker regions actually inside the ROI. This method is very fast and does not require the computational overhead of matching a circular shape through some polar transformation. As in Seoud et al. (2016), we define D as the diameter of the ROI, and the three following parameters:

- $d_1 = D/10$ is the estimated radius of the OD;
- $d_2 = D/360$ is the size of the smallest μA ;
- $d_3 = D/56$ is the estimated average size of the hemorrhages;

4.2.1.2 Illumination equalization

Similarly to Hoover and Goldbaum (2003), we overcome the effect of uneven illumination by estimating a local value of the background color and subtracting it channel wise:

$$I_{ie} = I + \mu - \mathbb{M}(I, d_1) \quad (4.1)$$

$\mathbb{M}(I, d_1)$ represents the median-filtered image where the median is locally computed in a window of size d_1 . The average intensity of each channel (computed inside the ROI) μ is added to keep the same color range.

4.2.1.3 Denoising

A simple denoising step is then applied to the three channels with a small Gaussian filter of size d_2 .

4.2.1.4 Contrast Limited Adaptive Histogram Equalization (CLAHE) on decorrelated channel

Histogram equalization tends to reinforce the contrast of an image by mapping its histogram to an uniform distribution. In practice, good results are obtained most of the time when this mapping is done locally on tiles (of size d_3). This is called Adaptive Histogram Equalization (AHE). However, AHE tends to amplify background noise in the fundus image. CLAHE overcomes this issue by clipping the original histogram at a predefined value. Moreover, applying CLAHE to the three channels independently might result in unrealistic colors. Instead, we transform the RGB image into the CIELAB color space. This space describes

the image as a combination of three components: its lightness \mathbf{L} and two color channels \mathbf{a} (green-red) and \mathbf{b} (yellow-blue). In order to preserve the color consistency, we only apply CLAHE to the lightness (\mathbf{L}) channel.

4.2.2 Data augmentation

It is a common practice to use data augmentation in deep learning in order to artificially augment the available training data. We use two types of simple augmentation: geometric and colorimetric. The first type affects both the input image and its corresponding ground truth segmentation, while the second type is only applied to the input.

The data augmentation is done on the fly during training. It consists in applying to each input image a random combination of one or more transformations among the following: rotation, shearing, elastic distortion, horizontal flipping, and scaling (for the geometry) and gamma, brightness, saturation, and contrast transform (for the colorimetry). Each individual transform is parameterized by a random amplitude (for example, the angle of rotation or the multiplicative coefficient for contrast), and re-sampled at each training iteration. With this strategy, the network is guaranteed to never see the exact same image twice.

4.2.3 Patches extraction

For lesions segmentation, the network is trained on square image patches. A fixed number of patches is extracted per image. The distribution of the patches' centers is conditioned by the ground truth: A prior favors patches centered on lesions (which helps counterbalancing the high imbalance of classes, the background being the predominant class). The prior is stronger on red lesions than bright ones, as they tend to be smaller and sparser. In images that do not contain any lesions, the patches' distribution follows a two dimensional uniform law. At inference time, however, the whole image is given to the network.

Each patch has a size of 256×256 pixels. This size represents an equilibrium between memory availability for the GPU and the need for context aggregation (i.e. the larger the patch is, the more contextual information the network can learn, assuming it is deep enough).

4.2.4 Multitask architecture

Our proposed neural net architecture is mostly fully convolutional. It is composed of three blocks illustrated in Figure 4.2. The first block is composed of an encoder, progressively increasing the number of features while decreasing their resolution. This lead to a higher receptive field at each successive convolution. The second block is composed of two decoders,

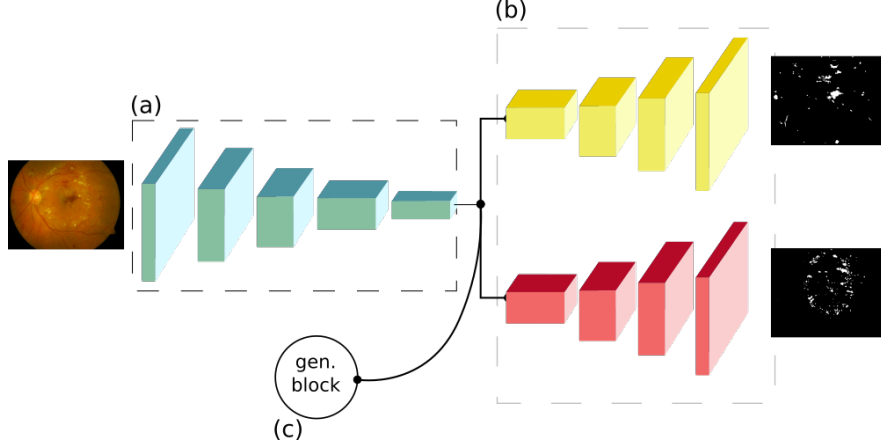


Figure 4.2 Simplified representation of the proposed architecture. (a) An encoder block is shared among two decoders (b), each of which is specialized in its class of lesions (red or bright). An additional block (c) is trained on detecting whether the presented patch is pathological or not. During the first step of the training, all branches are trained. During the second step, only (a) and (c) are trained on full images.

identical in their architecture and connected to the same encoder, but each having its own set of weights. This approach is called *Hard Parameter Sharing* Ruder (2017). The encoder/decoder ensemble is largely inspired by the U-Net, introduced in Ronneberger et al. (2015); nonetheless, we propose several modifications that are described below.

In addition to the aforementioned two blocks, a third block is also plugged into the encoder, which we call the *generalization block*. This branch performs binary classification to detect whether any sign of pathology is present or not in the given patch.

Due to memory constraints and the use of small mini-batches, no batch normalization is used. Instead, we choose the scaled exponential linear units (selu) activation function introduced in Klambauer et al. (2017) and its associated weights initialization.

Figure 4.3 shows a detailed overview of the whole architecture.

4.2.4.1 Encoder architecture

In addition to the selu activation function, we also amend the U-Net encoder with the following modifications:

- Mixed-pooling layer instead of max pooling. Mixed pooling is described in Lee et al. (2018) and is a compromise between max and average pooling. Various strategies are presented to combine both. Due to its simplicity and low memory overhead, we choose the *one parameter per feature* strategy. For each feature map \mathbb{X}_l , the mix is computed

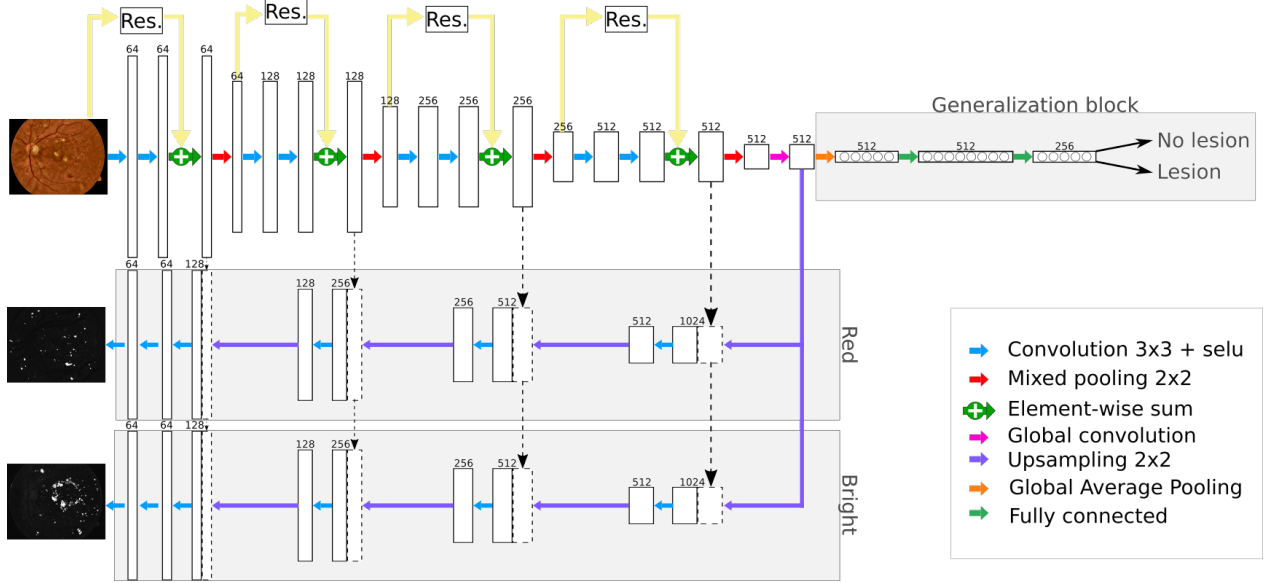


Figure 4.3 Multi-task architecture. The encoder is shared for the two decoders, trained to segment bright and red lesions, and the generalization block, trained to predict whether the input contains a lesion. By training the generalization block, we also train the encoder.

as:

$$f_{mixed}(\mathbb{X}_l) = a_l \cdot f_{avg}(\mathbb{X}_l) + (1 - a_l) \cdot f_{max}(\mathbb{X}_l) \quad (4.2)$$

where a_l is a parameter learned using classical backpropagation.

- As introduced in He et al. (2016), we use residual connections to facilitate the training. A residual connection consists of a 1×1 convolution with linear activation function that bypasses two "normal" convolutions, therefore allowing the network to exploit a simple identity mapping.
- At the lowest level of the encoder, we replace the 3×3 convolutions by global convolutions (GC) introduced in Peng et al. (2017). The idea behind GC is to exploit the large field of view offered by convolutions with large kernels without the computational and memory overheads. Figure 4.4 illustrates this principle.

4.2.4.2 Decoder architecture

The decoder architecture is the same as in the standard U-Net, except that we use the selu activation function. The decoder consists of successive upsampling layers followed by 3×3 convolutions. The upsampling layers are implemented as in Ronneberger et al. (2015). Each decoder is associated with its cost function. For both of them, we use a loss based on Cohen's coefficient κ . As for Dice-based losses, κ -loss is actually a soft (differentiable) approximation

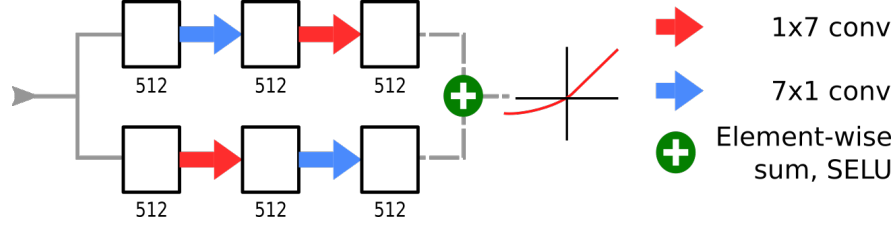


Figure 4.4 At the lowest point of the encoder, large kernel convolutions are simulated by the successions of a 7×1 convolution and a 1×7

of the real κ . Unlike maximum likelihood, the κ coefficient offer a chance correction and as lesion pixels only represent a tiny portion of the image, it advantageously corrects this class imbalance. κ is defined as the relative distance between the degree P_a of agreement between two observers (the ground truth and the prediction) and the probability of agreement P_r if the observations were randomly sampled:

$$\kappa = \frac{P_a - P_r}{1 - P_r} \quad (4.3)$$

During training, the κ -loss tends to be unstable over the first few iterations. To address this, we follow the recommendation given in Fauw (2015) and sum it with a clipped log-likelihood loss, which stabilizes and accelerated the convergence.

The last layer of a convolutional network is usually fed to a softmax activation function. But when using a κ -loss, post-softmax activations tend to be almost binary. Therefore, in order to obtain informative probability maps at inference time, we changed the softmax activation to a slightly modified function f defined as:

$$f(\mathbf{x}) = \text{softmax}(\alpha \mathbf{x}) \quad \text{with} \quad \alpha < 1 \quad (4.4)$$

This does not change the network's predictions, but produces more uniformly distributed probability maps, which we use to obtain smoother Receiver Operating Characteristic (ROC) curves for pixel-level performance evaluation.

4.2.4.3 Generalization block

The features maps from the encoder are fed into the *generalization block* through a global average pooling, which simply gives a spatial average of the different features maps. Global Average Pooling (GAP) reinforces the correlation between an entire feature map (and not just a local portion of it) and the prediction. Moreover, GAP is independent of the spatial

resolution of the input, which is an important property that we use to move from patch to whole image training as described in section 4.3.2. The rest of the generalization block consists of two hidden layers of fully connected neurons trained to do binary classification with a regular cross-entropy loss function.

4.3 Experiments

This section presents the experimental setup for training the proposed architecture as well as the validation process. We then assess the results of our methodology qualitatively and quantitatively on different independent databases. We evaluate its lesion segmentation performance using the standard metrics (sensitivity, specificity, precision, accuracy); as well, with screening applications in mind, we evaluate its image-wise performance (i.e. accurate detection of at least one lesion in a pathological image).

4.3.1 Material

We use the publicly available databases DIARETDB1 (Kauppi et al. (2007)), e-optha exudate (Decencière et al. (2013)), and EyePACS, as well as a private database for training and evaluation. DIARETDB1 contains 89 images, with red and bright lesions roughly annotated by multiple medical experts. We have refined the boundaries of these lesions under the supervision of an ophthalmologist. The e-optha exudates database consists of 47 images with exudates and 233 images with no lesions. We have manually annotated the red lesions in the 47 pathological images. In addition, the private database we had access to is composed of 19 images with red and bright lesions annotated by a medical expert. Finally, for image-level validation, we use the 1200 images of the Messidor database (Decencière et al. (2014)).

For the weakly supervised training phase, we take advantage of the large database provided by EyePACS for the Kaggle Diabetic Retinopathy Detection challenge. It contains over 35k images and their associated DR grade. Qualitatively, it is easily noticeable that a significant ratio of the images are of poor quality, exhibiting poor focus, very low illumination or containing multiples artifacts. But in a clinical situation, those artifacts are representative of the high variability in image quality that is often encountered.

We use data from various sources and different resolutions but the field of view of all the fundus images is commonly between 40° and 50° . Thus, we use the ROI diameter as a scale invariant, and each image is resized and cropped to be centered on the ROI, with a size of 1152×1500 pixels.

In order to ensure that the network does not overfit on a given database, we do cross-validation: for performance evaluation on e-optha, we only train on images from DIARETDB1

and the private database. Similarly, when we evaluate on DIARETDB1, only a subset of it is used for training (28 images), while the rest (61) is kept for testing. This division follows the instruction that come with the database. For each experiment, 15% of every training database is reserved for validation purposes. Over 150 epochs, the model with the highest κ score on the validation set is kept for performance testing.

4.3.2 Training

The training is organized in three phases, alternating between supervised and weakly supervised training.

Phase 1 The whole architecture is trained on patches. There are therefore three distinct costs (two from each decoder and one from the generalization block). The generalization block is trained to predict whether the patch contains a lesion of any sort: it is essentially a local abnormality detector. We hypothesize that this helps the network to focus on lesions rather than other salient markers and limits the number of false positives learned by the encoder. However, this training phase requires ground truth labels at the pixel level, which we only have in limited amount.

Phase 2 We propose a strategy to compensate this lack of data. The aim is to allow the network to generalize on more data without using pixel-wise training. We use the image data from EyePACS and consider that an image of DR grade 1 or higher contains at least one lesion (i.e. positive) whereas an image of grade 0 is assumed to be free of lesions (i.e. negative). Only the encoder and the generalization block are trained in this phase, now using entire images. Indeed, because the lesions position' in the image are unknown, we cannot have a patch-wise approach: in a positive image, there will be many negative patches and generally only a few positives ones. At this step, due to hardware constraint, images are resized to 576×750 pixels.

Phase 3 The generalized encoder is plugged back in the whole architecture and Phase 1 is repeated. Indeed, as the encoder's weights have been modified in Phase 2 but not the decoders', the latter require fine-tuning, and this is generally only a matter of a few epochs. We use mini-batches of size 8 in Phases 1 and 3 and of size 3 for Phase 2. We choose the Adam optimizer with a learning rate of 0.00001. We also use a weight decay $L_2 = 0.001$. It is important to note that some public databases provide only one sort of annotation (red or bright, not both). The strength of multi-tasking is to offer the possibility of fine-tuning a

branch independently from the others according to the availability of the labels.

4.3.3 Results

We evaluate three aspects of our system: the qualitative performance of the preprocessing method, the quality of the segmentations, and finally the capacity of the model to detect pathological images (in a screening context).

Preprocessing step We expect our preprocessing method to be able to enhance the lesions without increasing the signal to noise ratio. Moreover, the preprocessing must correct the overall image illumination. Figure 4.5 shows two examples of preprocessed images including one with a challenging problem of low contrast and very uneven illumination. As can be observed, the global color consistency is preserved in each image; this distinguishes our method from most of the previous methods proposed for fundus images.

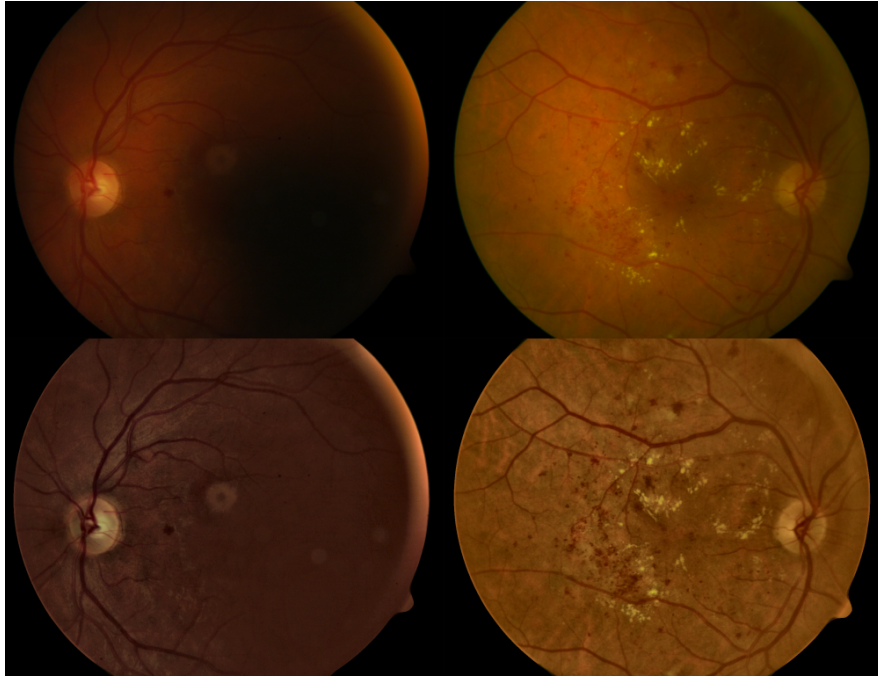


Figure 4.5 Color preserving preprocessing: the top row shows the original images, bottom row shows the results of the preprocessing step.

4.3.3.1 Lesion level validation

At a pixel level, we choose to evaluate the performance of the lesion segmentation task following the methodology proposed in Zhang et al. (2014). To summarize, rather than

simply counting pixels that are correctly classified, we count as true positive pixels those that belong to connected components that overlap with the ground truth by a minimal ratio. Conversely, if the components do not overlap by this ratio, the pixels will be counted as false positive. Moreover, if a connected component in the ground truth does not overlap with a corresponding lesion in the prediction by the same minimal ratio, the non-overlapping pixels will be counted as false negatives. For validation, it is difficult to do mutual comparison with state of the art works, as we propose a pixel-level segmentation of multiples lesions without discriminating between them. This justified the training of baseline models with which to compare the proposed one. We trained multiples U-Nets (one per lesions and per training database) following the original methodology detailed in Ronneberger et al. (2015). As the U-Net is a single-task architecture, this validation allows us to illustrate the potential benefits and drawbacks of multi-tasking. Note that we attempted to train a single U-Net with multiples classes as a target (red, bright, background) but we did not succeed in obtaining convergence with this model. We report our evaluation results in Tables 4.1 and 4.2 while Figures 4.6a and 4.6b show the ROC curves for the DIARETDB1 and e-ophta image sets.

Table 4.1 Comparative performance of the proposed approach for bright lesions segmentation

Method	Database	Images	Pixel-based results (%)			
			Sensitivity	Specificity	Precision	Accuracy
Baseline						
U-Net	e-ophta (EX)	280	66.57	99.78	19.35	99.76
	DIARETDB1	61	91.53	99.77	50.90	99.75
Multitask						
architecture	e-ophta (EX)	280	66.29	99.97	60.40	99.94
	DIARETDB1	61	78.35	99.89	64.16	99.85
Proposed						
architecture with genera- lization	e-ophta (EX)	280	70.36	99.97	66.47	99.95
	DIARETDB1	61	85.92	99.84	60.56	99.80

For both bright and red lesions and on both databases, in comparison with the multi-task architecture, we observe a clear improvement of the sensitivity with the training done on EyePACS, while keeping a similar precision. This is confirmed by the Area Under the Curve (AUC) in Figures 4.6a and 4.6b. However, on bright lesions we observe a lowering of the precision for DIARETDB1, which means an increased number of false positives. Visual

Table 4.2 Comparative performance of the proposed approach for red lesions segmentation

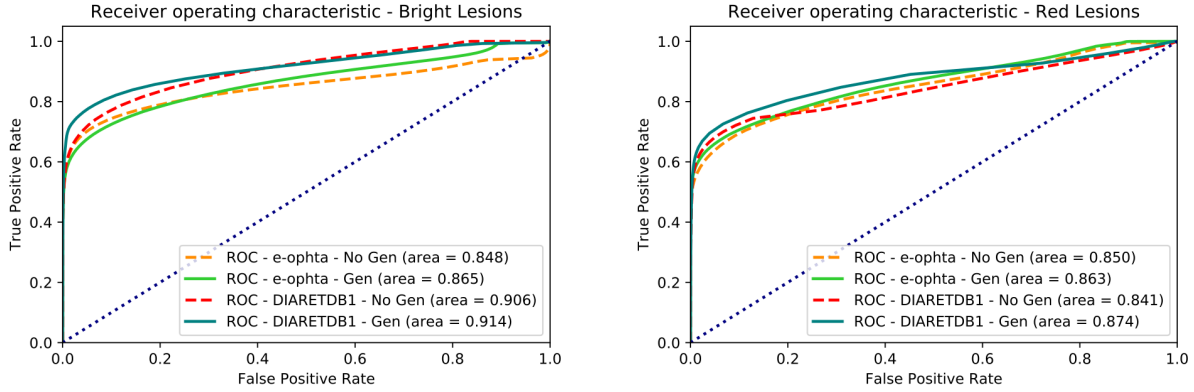
Method	Database	Images	Pixel-based results (%)			
			Sensitivity	Specificity	Precision	Accuracy
Baseline						
U-Net	e-ophta (EX)	47	76.46	99.91	68.90	99.86
	DIARETDB1	61	84.54	99.79	57.94	99.73
Multitask						
architecture	e-ophta (EX)	47	69.50	99.94	75.13	99.87
	DIARETDB1	61	73.03	99.95	82.20	99.87
Proposed						
architecture with genera- lization	e-ophta (EX)	47	75.05	99.94	73.06	99.88
	DIARETDB1	61	73.90	99.95	83.59	99.87

inspection of the probability maps before and after generalization provides some insight into the differences between the two multi-task architectures:

- The larger number of false positives in the generalized model comes from a few images where the optic disc is partially detected as a lesion. This also explains the lower AUCs in the per-image validation in Figure 4.7a.
- However, the generalized model is better at detecting non hard-exudate lesions; in particular, it yields more precise segmentations of lesions with blurred borders. One possible interpretation is that the non-generalized model, trained on e-ophta, had mostly hard exudate lesions to learn from. Others lesion types were brought by the weakly supervised training phase using EyePACS.
- In both models, we observe that the majority of false positives come from nerve fibers classified as lesions. In fact, this is probably the main shortcoming of our model, as the presence or absence nerve fibers in fundus images is not (in most cases) diagnostically informative.

4.3.3.2 Image level validation

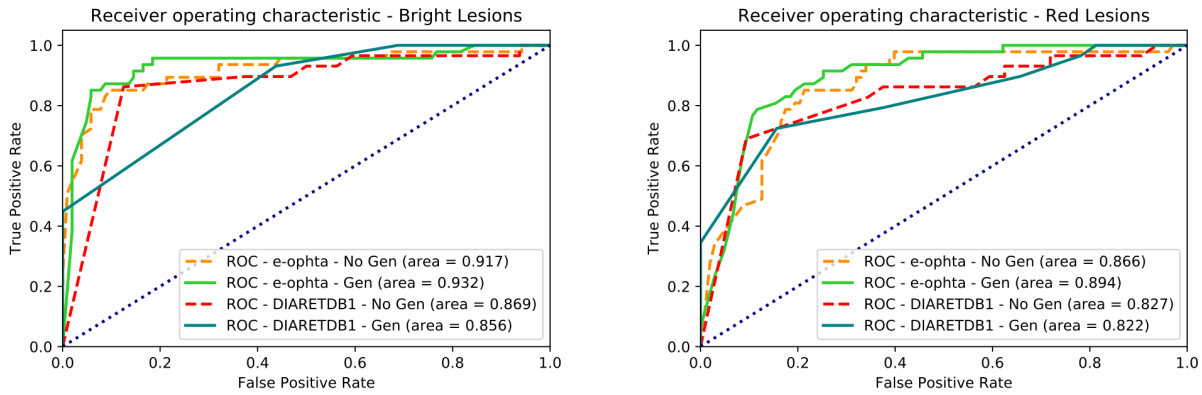
In addition to the evaluation at the pixel level, it is also informative to check the system’s ability to predict whether an image is referable to a clinician, the criterion being the presence of at least one lesion (one positive pixel). In other words, the aim here is to evaluate if the algorithm could be used as is in a screening program. By varying the cutoff threshold on



(a) Comparative results for bright lesions segmentation. (b) Comparative results for red lesions segmentation.

Figure 4.6 Validation at a pixel level of the model on the two databases DIARETDB1 and e-ophta. We show the result before and after the generalization step done on EyePACS. In all configurations, this step improves the segmentation performance.

the probability maps, we can plot the screening ROC curves for both red and bright lesions separately, as shown in Figures 4.7a and 4.7b. In table 4.3, we propose a comparative study with previous works and the baseline method. Our performances are much higher than those of the baseline (U-Net) and quite close to those of the compared methods.



(a) ROC for pathological image detection based on bright lesions. (b) ROC for pathological image detection based on red lesions.

Figure 4.7 Validation at an image level of the model on the two databases DIARETDB1 and e-ophta. A prediction containing at least one positive pixel is considered as referable.

Table 4.3 Comparative performance on detection at an image-level.

Detection Method	Database	Lesions	AUC
Seoud et al. (2016)	Messidor (R0 vs All)	Red	0.899
van Grinsven et al. (2016)	Multiples	Bright	0.883
Baseline model			
U-Net	e-ophta (EX)	Bright	0.56
	DIARETDB1	Bright	0.73
	e-ophta (EX)	Red	0.74
	DIARETDB1	Red	0.80
Proposed architecture	e-ophta (EX)	Bright	0.932
	DIARETDB1	Bright	0.856
	e-ophta (EX)	Red	0.894
	DIARETDB1	Red	0.822
	Messidor (R0 vs All)	Red	0.837
	Messidor (R0 vs All)	Bright	0.696

4.4 Discussion

This work proposes several contributions, from a new image preprocessing strategy to a novel method for taking advantage of weak supervision for lesions segmentation. While the preprocessing represents only an incremental improvement over previous methods, it differs from them by the fact that it preserves color consistency within images without trying to standardize the colors between images. This latter case typically consists in color transfer from a target image to a source image (considered of good quality) or to a given distribution. It can be done globally at low computational cost by assuming that the pixels' distribution per decorrelated channel is Gaussian. Nonetheless, this strategies may locally induce non-realistic colors. Alternatives usually rely on local histogram matching, which provides higher fidelity but to the detriment of computational efficiency Schmitt (2017).

In our preprocessing method, we intentionally avoid this step. Indeed, we argue that the color diversity present in the fundus dataset might actually contain diagnostically relevant information. Indeed, even if it has been shown Sisson et al. (2014) that color distribution is highly dependent of the hardware used for the acquisition, it also depends on the age of the patients and may even be an indication of hyperglycaemia exposure in diabetic patients M. et al.. Therefore, as it might be diagnostically relevant, we don't change the histogram of distributed colors.

Concerning lesions segmentation, there have been many works proposed to do automatic diagnosis from a fundus acquisition; these generally follow one of two schemes:

- A lesion based approach, consisting in detecting and/or segmenting some type(s) of abnormal biomarkers. In this scenario, when a training strategy is deployed, the goal will be to minimize a pixel-wise cost. It relies therefore on manually annotated images of the targeted class at a pixel level, which is a rare resource. The result of the detection/segmentation can be fed into a classifier to obtain a disease-associated grade.
- A global approach, fostered by the emergence of deep learning algorithms, that can be summarized as data-mining in a generally large database in order to let the algorithm learn by itself the features that can distinguish between each disease grade. In this configuration, only labels at an image level are needed, but consequently, the system is not transparent in its decision making process.

For diabetic retinopathy screening and grading, public databases are available for both approaches. Nonetheless, for pixel-level prediction, labelled data are usually coarse and/or sparse, which raises the question of the generalization capability on new data. By contrast, the global approach can benefit from large existing databases, but we argue that the "black box" nature of the decision-making in such systems is not acceptable from a clinical standpoint. Our method takes advantages of both approaches to improve the performance on lesions segmentation.

The question of the comparison with current state of the art methods naturally arises. It appears that the lack of standardize procedure makes the comparison quite difficult. In Kaur and Mittal (2018), authors report a performance of 83.00%-89.40% (sensitivity-precision) on a subset of 82 images of e-ophta for exudates segmentation. While those are indeed higher performance than ours, there is no indication on how was chosen this subset. Moreover, they report a performance of 71.34%-79.89% on the whole DIARETDB1, which is above ours but at the cost of a largely lower specificity (88.56%) and accuracy (85.34%). In Zhang et al. (2014), authors obtain a performance of (74.00%-72.40%) on DIARETDB1, but don't report their performance on another database. In our experiments, our system was prone to more false positives than other methods. To explain this, we consider that, in the case of bright lesions, the network was trained for the wider task of lesions segmentation, which includes drusen and cotton wool-spots, rather than simply exudates. For red lesions, most of the literature focuses on μA detection rather than full red lesions segmentation. To our knowledge, there is no existing work with segmentation performance reported at a pixel level to which we can compare ourselves. This motivated the choice of an image-level validation procedure in addition to the pixel-level tests.

We have shown that the generalization procedure proposed significantly improves the performance of our multi-task architecture in term of precision and sensitivity on different databases. In Figure 4.8, we present visual examples of lesion segmentation in images presenting different levels of difficulty. These examples illustrate how our model is relatively robust to various artefacts, uneven illumination and fuzzy (out-of-focus) images. As pertains to the limitations of the proposed model, the major one is its tendency to predict nerve fibers as bright lesions, as previously mentioned. In some cases, we also observed false positive detections caused by photocoagulation marks resulting from laser treatment. As is often the case, the availability of more training data should remove those false positives. Moreover, as photocoagulation marks are very uniform in their shape and color, they could also be removed without machine learning by using morphological operations or matched-filtering. Concerning red lesions, we found some cases of very large haemorrhages for which the network failed to predict the true boundaries and fragmented the lesions into multiple smaller ones (see Figure 4.8b). This indicates that the network lacks a receptive field large enough to take into account the wider image context. A multi-scale segmentation framework could correct this issue.

4.5 Conclusion

The proposed approach provides a fully trainable model for joint segmentation of bright and red lesions in retinal images and proposes a novel methodology for exploiting image-level labels to improve segmentation performance. Thanks to the intensive work on optimization of deep learning computation for GPUs in the last few years, a complete segmentation using our system takes less than a second with a NVIDIA GTX 1070Ti running on a standard desktop PC. This fast execution time makes the model suitable for clinical deployment. In particular, future work will focus on assessing how it can accelerate the process of labelling data by a clinician, by providing lesion pre-annotations. Following steps will include classifying each segmented lesions. This will serve two purposes: building a detailed atlas of the retinal lesions and providing an automatic diagnosis built upon a set of rules based on the detected lesions. Both of these developments will contribute to a computer-aided diagnosis system that is transparent in its decision-making.

4.6 Acknowledgement

The authors gratefully acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for its support and Philippe Debanné for revising this manuscript.

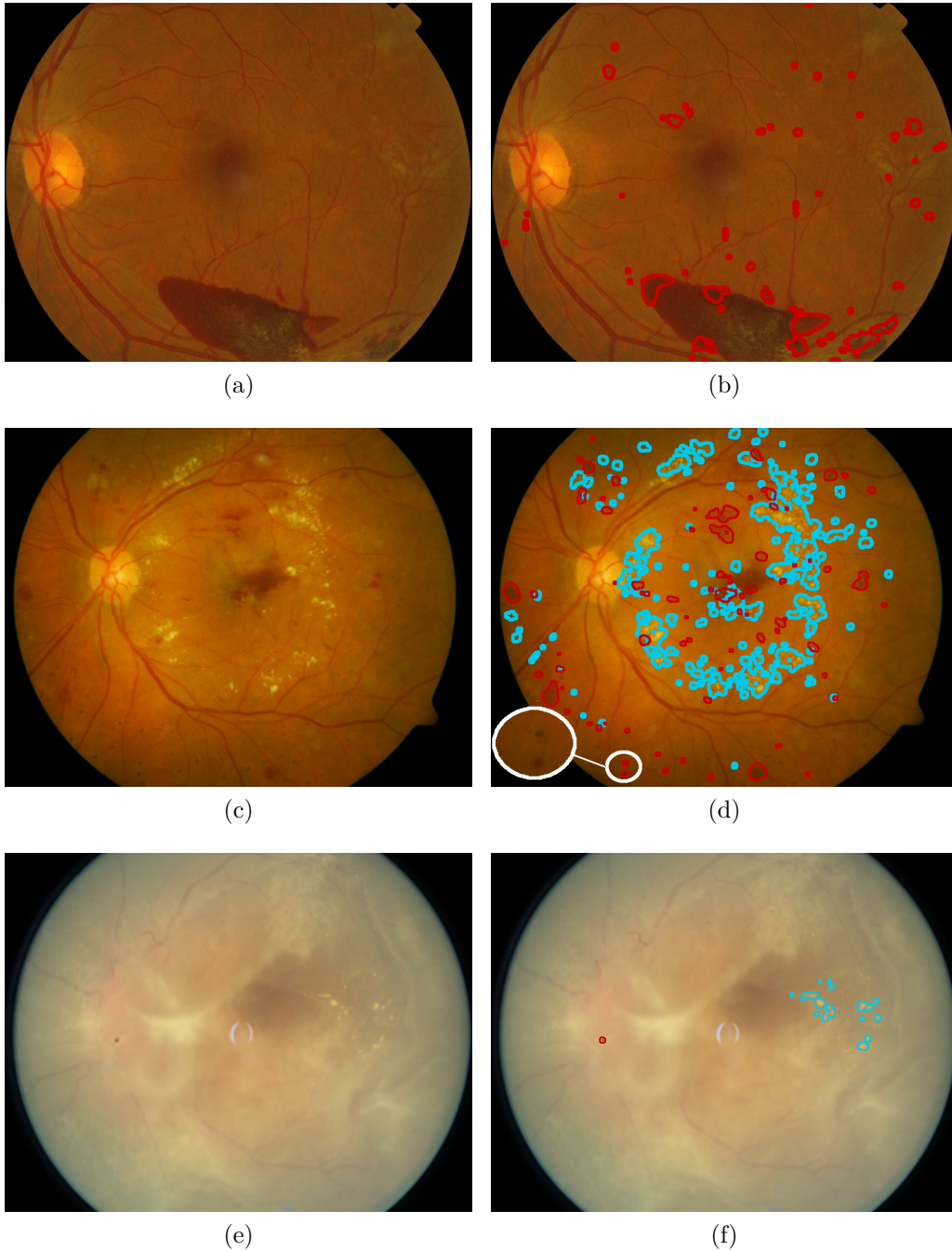


Figure 4.8 Lesion segmentation examples. Figures (a), (c) and (e) are the original images extracted respectively from Messidor, DIARETBD1 and EyePACS. Figure (b) only shows the red lesions detection: it illustrates the failure to segment the large haemorrhage at the bottom as a single region. Some small vessels are also counted as false positives in the vicinity of the OD. In figure (d), two biomarkers detected by our model as red lesions are highlighted. The bottom one is indeed a μA but the top one is a false positive photocoagulation laser mark. despite this specific case, only a few of the many laser marks in this image are classified as lesions, proving that this is not a systematic misclassification. Figure (f) illustrates a successful case of exudate segmentation in a poor quality, very low-contrasted image. Many other approaches would require a prior segmentation of the optic disk while it hardly visible in this case.

CHAPITRE 5 DÉVELOPPEMENT D'UNE PLATEFORME D'ANNOTATIONS

Le travail de segmentation automatique des lésions a été rendu possible grâce à diverses bases de données accessibles publiquement. Pour certaines, il a cependant fallu compléter les annotations cliniques pour détourner précisément les frontières des lésions. Cette étape, bien qu'indispensable à l'entraînement de l'algorithme, est longue et fastidieuse. Par ailleurs, elle requiert une grande expertise médicale et doit donc être confiée à des médecins qui n'ont pas nécessairement beaucoup de temps à consacrer à la tâche. Et pourtant, comme le montre le tableau 2.1, le besoin existe. La quantité de données accessibles reste relativement marginale et ne couvre pas l'ensemble des cas de figure susceptibles d'être rencontrés en situation clinique. Face à ce constat, une équipe a été mise en place, réunissant des spécialistes de plusieurs domaines incluant des ophtalmologistes de Montréal, Sherbrooke, Vancouver et Toronto. La décision a été prise de concevoir une plateforme de récupération de données permettant l'annotation par des experts afin de constituer une base de données de référence capable d'entraîner des algorithmes de segmentation et/ou de classification. Une contrainte est imposée sur les algorithmes de diagnostic : ils doivent être transparents dans leur prise de décisions. Ainsi par exemple, un dépistage positif à la rétinopathie diabétique doit être justifié par un marquage des lésions qui ont conduit à cette décision. Cela exclut les solutions globales présentées à la section 2.3, car celles-ci ne s'appuient pas sur des caractéristiques intelligibles pour établir leur diagnostic.

Par ailleurs, l'algorithme de dépistage, même s'il se concentre sur la rétinopathie, doit pouvoir repérer toute image présentant une anomalie : d'un point de vue clinique, il pourrait être dangereux de laisser croire à un patient qu'il n'est pas malade sous prétexte que le système automatique qui l'a traité est incapable de détecter la pathologie dont il est atteint.

Ce chapitre présente les contraintes spécifiques au problème de l'annotation d'images médicales et les outils qui ont été imaginés pour y répondre. Dans la mesure où le développement de la plateforme n'en est qu'à ses prémices, l'accent est mis sur l'aspect conceptuel qui servira de guide au développement ; néanmoins, le chapitre se conclut par une présentation des solutions technologiques adoptées et du prototype dans son état d'avancement actuel.

5.1 Spécificités de l'annotation d'images de fond d'œil

L'annotation du fond d'œil consiste au tracé d'un masque de segmentation sémantique (c'est à dire le marquage des différentes régions de l'image, associées à leur sens). Ces différentes

régions sont de taille et de forme très variées. A titre d'exemple, un micro-anévrisme ne dépasse pas quelques pixels de large, là où les vaisseaux forment un arbre vasculaire couvrant l'ensemble de l'image. La difficulté pour le clinicien est, dans le premier cas, de repérer la pathologie, dans le second cas, de marquer précisément l'épaisseur et l'identification en tant que veine ou artère de chaque vaisseau (importante pour le calcul de l'AVR). Une série d'essai a été menée pour déterminer le temps d'annotations des lésions et du réseau vasculaire sur plusieurs images et une estimation du temps moyen d'annotations en fonction du marqueur considéré en a été tirée (tableau 5.1). Bien que très approximatives, ces valeurs permettent de borner le temps d'annotations par images. En incluant les biomarqueurs non pris en compte dans ce tableau (macula/fovéa et DO incluant excavation papillaire), on peut raisonnablement estimer à 30 min le temps d'annotations moyen d'une image entière. A ce chiffre peuvent être ajoutées quelques minutes pour le chargement du fichier et l'évaluation de la qualité de l'image.

Cependant, de toutes évidences, de grandes fluctuations sont à attendre. L'annotation d'une image saine est bien plus rapide que celle d'une image comportant de nombreuses lésions. Par ailleurs, tous les cliniciens n'ont pas la même aisance avec les outils informatiques et certaines lésions très petites peuvent être ratées par certains. Ces essais ont été réalisées avec l'outils

Tableau 5.1 Temps d'annotations (moyenne/écart-type) par marqueurs.

Type	μ_{image} (min :sec)	σ (min :sec)	Nombre d'images.
Lésions (observateur 1)	7 :40	-	5
Lésions (observateur 2)	5 :20	1 :45	5
Veines/Artères (observateur 3)	12 :20	2 :25	3

open source The Gimp¹ configuré pour la tâche. Cependant, étant un logiciel généraliste, un certain nombre d'opérations ne sont pas optimales pour une application spécifique telle que l'annotations. Les temps relevés dans nos expériences représentent donc plutôt une fourchette haute.

Différentes stratégies ont été proposées pour réduire au maximum ces temps d'annotations. La principale serait de fournir aux cliniciens une pré-annotation automatique de l'image, provenant des différents algorithmes conçues en interne au LIV4D, incluant l'algorithme de segmentation de lésions présenté au chapitre 4. La tâche du clinicien se limiterait dans le meilleur des cas à une validation de la segmentation proposée ou une simple correction des erreurs si nécessaire.

Cette approche devra faire l'approche d'un processus de vérification des résultats. Il est

1. <https://www.gimp.org/fr/>

possible la pré-annotation fournie par l'algorithme introduise un biais dans le jugement du clinicien chargé de l'annotation. Un protocole simple de type « double-aveugle » permettra de confirmer ou non l'existence de ce biais. Une approche possible à trois experts serait :

- Obtenir une segmentation complète de l'expert 1 en partant de l'image brute.
- Obtenir une segmentation complète de l'expert 2 en partant des pré-annotations automatiques.
- En cas de désaccord entre les deux, proposer à l'expert 3 trancher. Il faudra évaluer si la source du désaccord provient de pré-annotation automatique et si oui de quel côté tend à se ranger l'expert 3.

Pour tirer partie des pré-annotations automatiques, il est nécessaire que l'interface développée propose des outils de visualisation d'images efficaces, d'outils d'édition accélérant la correction et qu'elle limite au maximum les temps de traitement entre deux images. La section suivante présente les solutions imaginées pour répondre à ces contraintes.

5.2 Conception d'outils dédiés aux cliniciens

La définition de l'interface s'est faite en préalable au travail de programmation, suite à des échanges avec les cliniciens. Différents outils de dessins ont été discutés pour optimiser leur rendement, incluant des outils de dessin dynamiques (ancrage du pinceau sous les lignes de contours de l'image, forme géométrique paramétrique), y compris vectoriels. La Figure 5.1 représente le visuel d'interface graphique retenu. L'organisation est structurée autour de trois panneaux verticaux. Le panneau central est dédié à la visualisation de l'image, incluant la possibilité de détacher des fenêtres flottantes proposant une visualisation à différentes échelles. Le panneau de gauche regroupe les outils de dessin, celui de droite les outils de sélections de calques. La barre horizontale supérieure permet de choisir entre les modules existants : exploration de la base de donnée, dessin, évaluation de qualité et paramètres utilisateurs. La notion de calques réfèrent en réalité aux différentes cartes de segmentations sont regroupées en catégories sélectionnables indépendamment (Figure 5.2). L'interface proposera également un module d'évaluation de la qualité de l'image par zones, illustré sur la Figure 5.3.

5.3 Collecte des données

La Figure 5.4 résume le fonctionnement de la base de données, en mettant l'accent sur l'aspect itératif de sa construction. Les images brutes, stockées sur un serveur centralisé, sont traitées par les algorithmes de rehaussement de contraste et de correction illumination. Parallèlement, les outils de segmentation génèrent les cartes de pré-annotations. Le tout est envoyé chez le

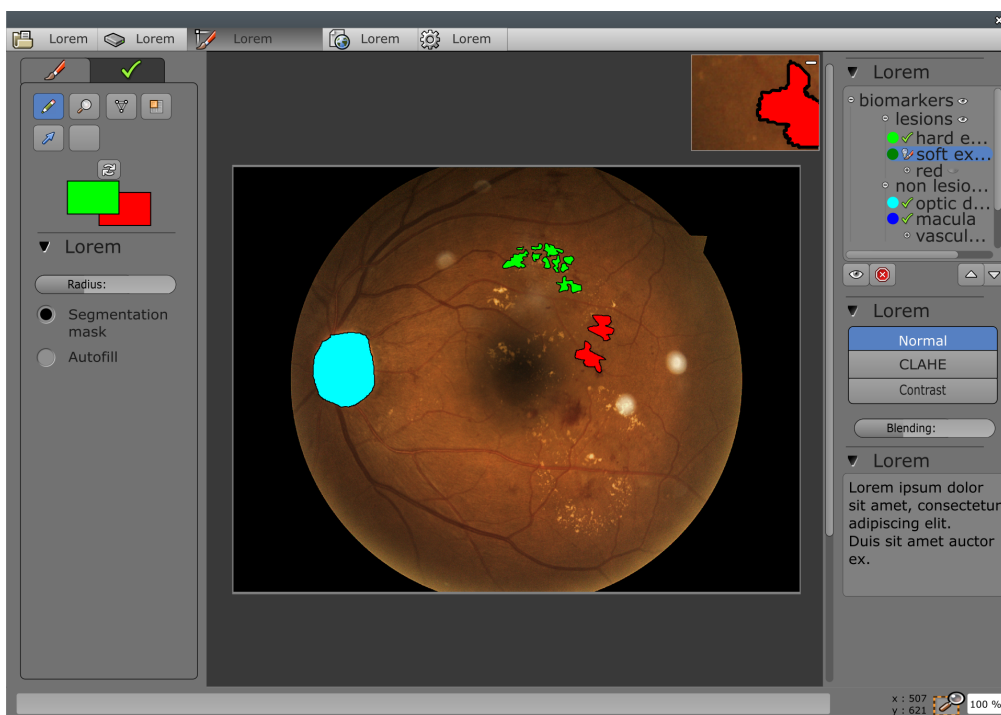


Figure 5.1 Proposition d'interface utilisateur graphique.

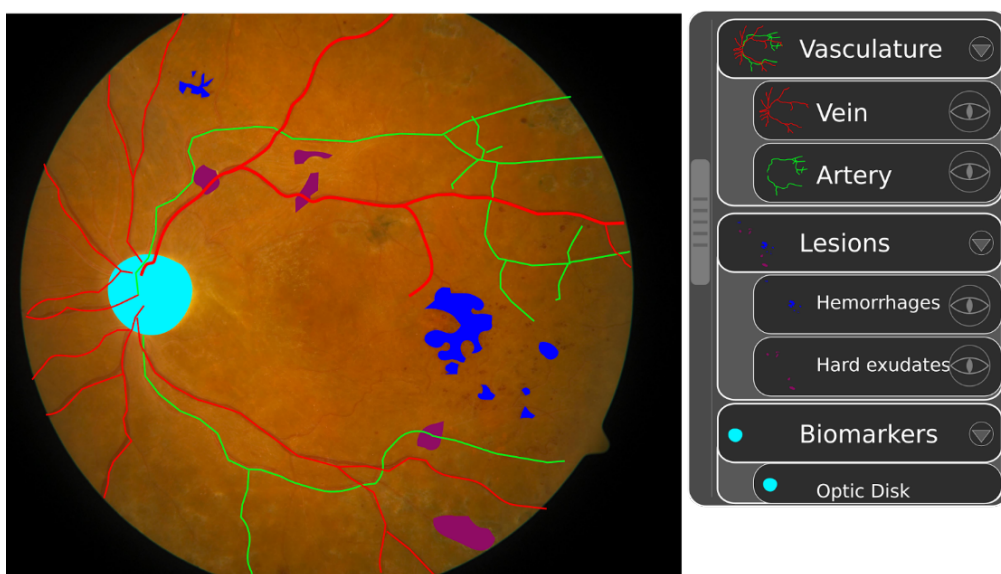


Figure 5.2 Organisation des annotations en catégories hiérarchisées.

clinicien participant au processus d'annotations. L'interface lui permet de valider ou corriger le cas échéant les pré-annotations. Les corrections sont renvoyées vers le serveur central et peuvent servir à améliorer les algorithmes de prédictions.

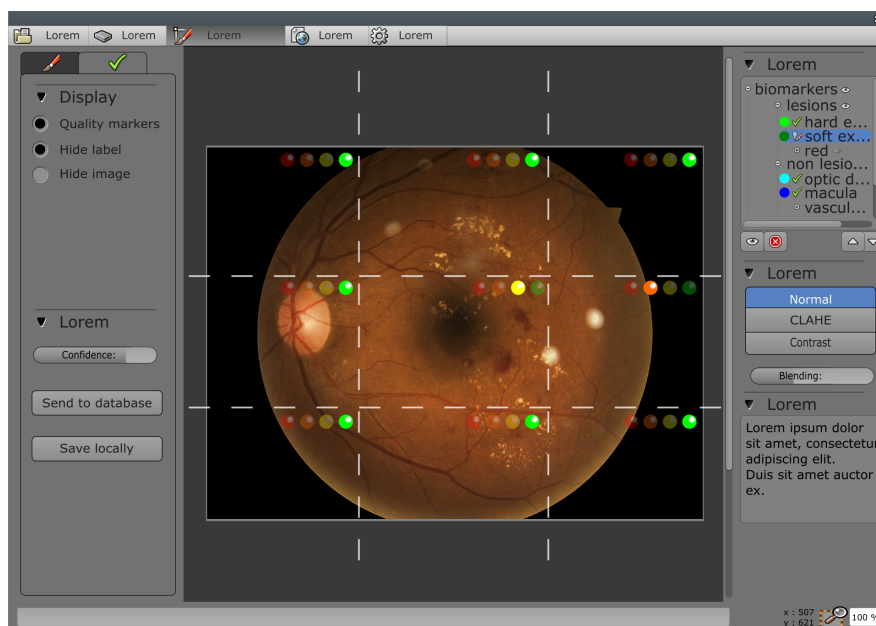


Figure 5.3 Interface d'évaluation de la qualité des images.

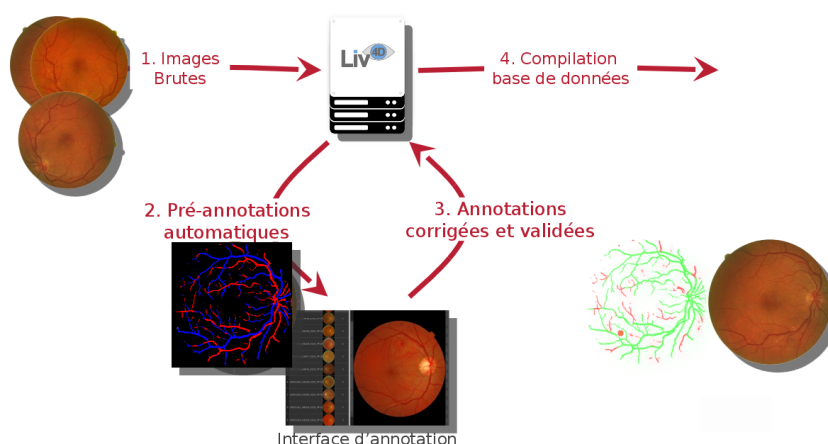


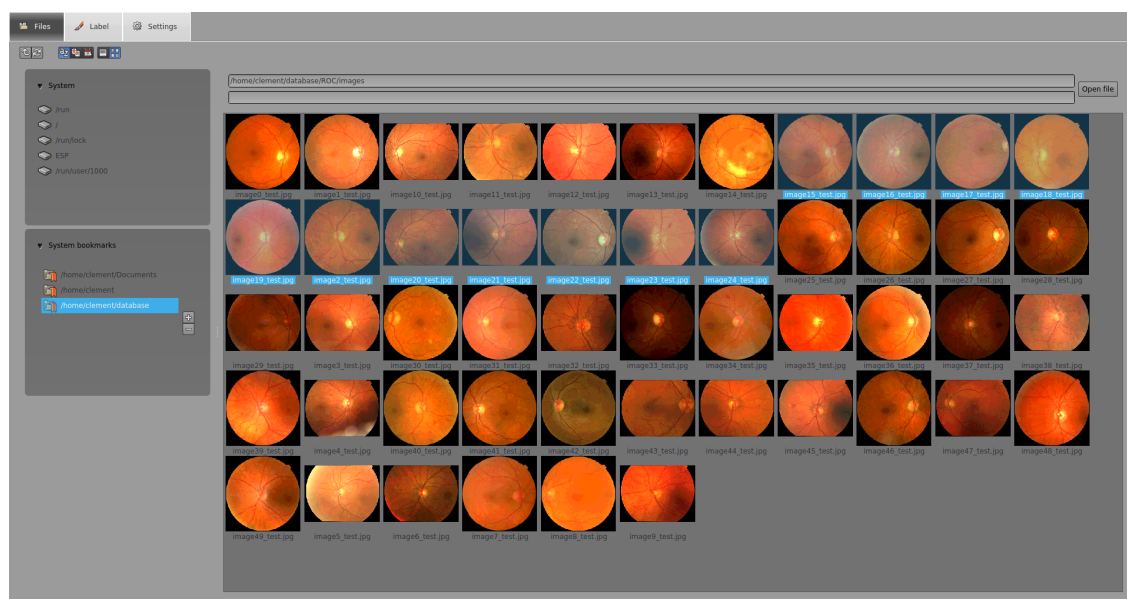
Figure 5.4 Organigramme de fonctionnement de la base de données proposées.

Par ailleurs, le serveur peut également être responsable d'une répartition « intelligente » des données à annoter, permettant notamment de faire annoter plusieurs fois certaines images par différents cliniciens pour étudier la variabilité inter-experts. Cette approche permettra de créer une vérité terrain plus fiable (issue de l'accord de plusieurs experts), mais également d'identifier les images présentant le plus haut degré de difficultés. Celles-ci pourront faire l'objet de plus d'attention lors de l'évaluation des performances des modèles automatiques. Il est également envisagé d'intégrer un module de communication dans l'interface permettant aux experts d'échanger leurs avis sur des cas problématiques.

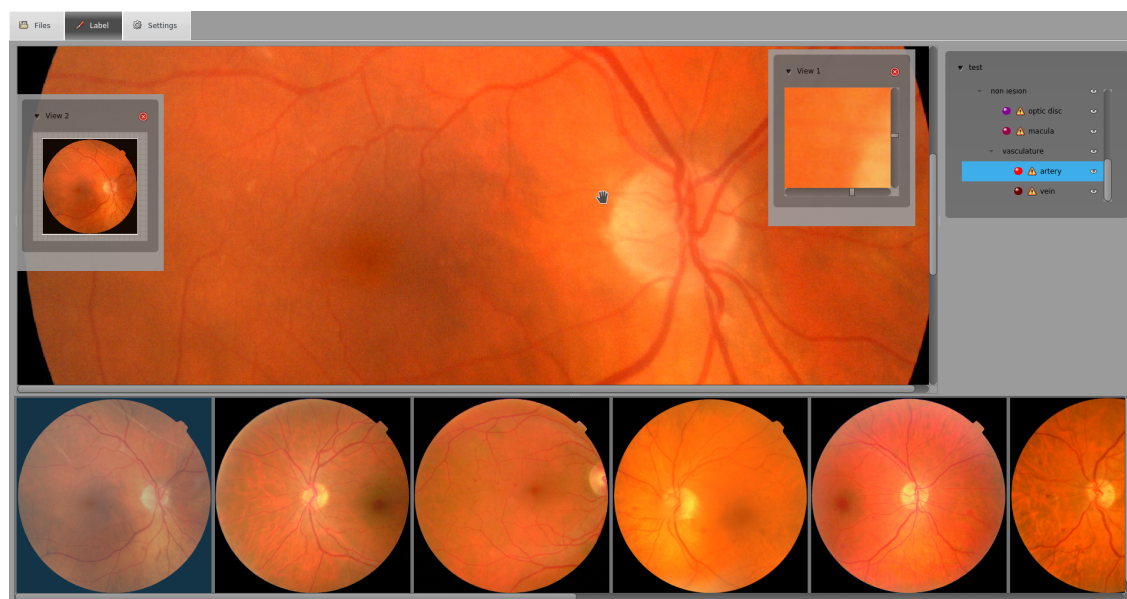
5.4 Choix de programmation

Au moment de la rédaction de ce mémoire, les différents visuels présentés de l'interface ne sont pas encore fonctionnels et seule une version précoce de l'interface a été développée. Néanmoins, un certain nombre de choix de programmation ont déjà été adoptés. Parmi ceux-ci, l'architecture de la base de données SQL a été déterminée, permettant de stocker des données issues de différentes modalités ainsi que des informations non confidentielles sur le patient. Pour l'interface, afin de proposer une solution multi-plateformes et performante, le choix s'est porté vers le langage C++ et la librairie QtWidget. Dans une certaine mesure, cela laissera également le choix d'un déploiement sur tablette ultérieurement, même si l'application reste pour l'instant développée pour ordinateur.

Le format d'enregistrement des fichiers dans la base (image et les différentes cartes d'annotations) est le SVG. Cette structure XML d'objets graphiques contient déjà un système d'organisation en groupes hiérarchisés des objets, équivalent au système de calques souhaité. La Figure 5.5 présentent des captures d'écran du prototype en son état actuel.



(a) Exploration dans la base de données



(b) Inspection d'une image dans l'interface. les fenêtres flottantes permettent d'avoir simultanément plusieurs niveaux d'agrandissement

Figure 5.5 Captures d'écran de l'interface dans son état actuel.

CHAPITRE 6 DISCUSSION GÉNÉRALE

6.1 Synthèse des travaux

Le travail présenté dans ce mémoire se compose donc de deux volets : la recherche en tant que telle s’est concentrée sur une nouvelle méthodologie de segmentation et en complément le développement d’une plateforme à destination des cliniciens a été initié. Cette plateforme permettra de créer la jonction entre des travaux de recherche, trop confinés au domaine de vision par ordinateur, et le déploiement clinique. Cette approche, particulièrement avec l’émergence de l’apprentissage profond, reste encore rare. Le foisonnement des travaux sur l’IA dédiée au biomédical, bien qu’extrêmement prometteurs par ailleurs, ne rend pas compte du déploiement effectif des techniques développées et de leur réel impact sur la santé. Pire encore, l’évaluation de la performance des algorithmes est parfois biaisée du fait de bases de données de test trop petites, trop semblables aux données d’entraînement ou ne reflétant que partiellement la réalité clinique. Ce biais peut amener à des conclusions trop optimistes sur la capacité réelle des algorithmes. Ce constat a fait l’objet d’un éditorial paru dans la prestigieuse revue *Nature* (Nat (2018)), soulignant ce problème et proposant des solutions. La teneur de celles-ci peut se résumer en une ligne : pour conclure de l’efficacité réelle d’un algorithme dédié au médical, une collaboration inter-domaines doit s’opérer entre chercheurs pour mener des études cliniques approfondies et rigoureuses. Il n’en reste pas moins que l’apprentissage machine (et les réseaux de neurones par excellence) semble particulièrement adaptée à cette tâche, grâce à la capacité de ces systèmes à généraliser de façon efficace et à s’améliorer au fur et à mesure de la construction de la base de données d’apprentissage (apprentissage par renforcement).

Le projet mené au cours de cette maîtrise vise donc à satisfaire ces deux pans de la recherche médicale (recherche en IA et validation sur le terrain auprès de cliniciens). Pour le premier aspect, une nouvelle méthodologie de segmentations des lésions est proposée. Pour la première fois, les deux types de lésions sont segmentés conjointement. Pour cela, une architecture de réseaux de neurones inédite a été proposée qui explore la segmentation multitâche. La comparaison des performances avec une approche monotâche encourage ce choix. Pour expliquer cette observation, l’hypothèse suivante est émise : l’apprentissage sur une tâche est en réalité avantageux pour la seconde si les deux sont corrélées. Dans le cas de la segmentation de lésions, la présence d’un type de lésions (par exemple sombres) n’est pas décorrélée de la présence de l’autre type (blanche), la détection d’un type de lésion conditionnant implicitement la probabilité de présence d’une seconde. En particulier, les caractéristiques extraites

par l'encodeur du réseau se concentrent sur la distinction entre tissus pathologiques et sains (et non pas entre lésions rouges -ou blanches- versus tout le reste).

De plus, l'architecture converge avec un nombre d'images annotées restreint (au mieux une centaine, quelque soit la configuration expérimentée) et obtient des résultats comparables avec l'état de l'art, tout en proposant une généralisation raisonnablement efficace sur d'autres bases de données.

En lien avec les avantages des architectures multitâches, l'apprentissage faiblement supervisé a été expérimenté, en entraînant l'encodeur de l'architecture sur des images annotées globalement (et non au niveau pixelique). Une autre base de données, beaucoup plus grande et composée d'annotations de gradation de la rétinopathie diabétique, a été exploitée pour cela. Par conséquent, l'encodeur du réseau a pu « voir » un nombre bien plus important d'images. Les performances en segmentation s'en sont trouvées améliorées, témoignant de l'intérêt de cette démarche. Cette méthodologie peut se répliquer sur des nombreux problèmes de segmentation, notamment médicale : la connaissance de la présence d'une pathologie dans une image est une information utile et exploitable pour la segmenter. L'approche proposée permet de mettre en pratique cette idée très intuitive.

L'entraînement est fait sur des images pré-traitées pour faire ressortir les structures de la rétine par rapport au fond. L'algorithme utilisé est une amélioration d'une méthode existante ; il a cependant l'avantage de préserver la cohérence colorimétrique de l'image d'entrée.

Enfin, pour contrebalancer le déséquilibre de classe inhérent à la segmentation de petits objets, une nouvelle fonction de coût basée sur le coefficient Kappa (κ) de Cohen est proposée. La démonstration analytique de l'efficacité de ce coefficient à prendre en compte ce déséquilibre dans le cas de la segmentation est donnée dans ce mémoire.

Du point de vue du développement logiciel de la plateforme d'annotations, un concept a été discuté et proposé aux différents acteurs du projet (qui dépasse du cadre de la maîtrise). Une fois le concept approuvé, le développement d'un prototype a démarré. La combinaison de différents algorithmes de segmentation devrait accélérer significativement le travail d'annotations mené par les cliniciens et permettre ainsi de créer une base de données de référence comportant le plus d'annotations expertes possibles. La constitution de cette base permettra alors d'améliorer d'autant plus les performances des algorithmes, grâce à l'apprentissage par renforcement.

6.2 Limitations de la solution proposée

Les différents points positifs de l'architecture proposée n'occultent pas les difficultés qu'elle peut éprouver à gérer certaines situations. Malgré des performances globales prometteuses,

la segmentation n'est généralement pas bonne dans les cas de figures suivants :

- Les gros objets (en général, des hémorragies faisant au moins la taille du disque optique) ont tendance à être fragmentés en plusieurs régions. D'un point de vue diagnostique, la conséquence n'est guère dommageable. En revanche, elle traduit la difficulté de l'architecture à gérer le contexte de l'image sur une longue distance. Ceci est symptomatique d'un champs de vision effectif du réseau trop restreint.
- Les fibres nerveuses sont trop systématiquement prises pour des lésions blanches, ce qui limite grandement l'applicabilité de la méthode à des fins de dépistage. En effet, les fibres nerveuses apparentes ne sont pas pathologiques.
- Dans certaines situations, d'autres marqueurs non pathologiques sont pris pour des lésions. En général, il s'agit de situations exceptionnelles : par exemple, des traces de photocoagulation laser peuvent être prises pour des micro-anévrismes, leur formes étant très proches.

Concernant la plateforme d'annotations, à ce jour, sa conception se focalise essentiellement sur l'annotations d'image de fond d'œil, bien que l'annotation d'image OCT a été envisagée. De façon générale, une des limitations de la plateforme est de ne pouvoir gérer qu'une seule image à la fois ; or de l'avis des cliniciens, il n'est pas rare de comparer plusieurs acquisitions du même œil, notamment si la première acquisition présente des zones d'ombres sous un certain angle.

Cependant la gestion de multiples acquisitions et de multiples modalités est un objectif secondaire, qui sera considéré une fois les outils d'annotations et de visualisation mis en place.

6.3 Améliorations futures

Pour améliorer les performances de segmentation, une piste évidente serait d'augmenter la quantité d'images d'entraînement. Cela n'est cependant pas possible en l'état des bases de données disponibles, bien que la mise en place de notre base de données devrait le permettre. Cependant, à plusieurs égards, il devrait être possible d'améliorer les capacités de l'architecture sans exploiter d'avantages d'images. Nous proposons ici des suggestions d'améliorations. Pour compenser le défaut du champs de vision trop restreint du réseau, une possibilité est de s'inspirer des travaux portant sur les modèles d'attention. Ceux-ci sont développés pour permettre une prise en compte du contexte de l'entrée à longue portée. Différentes solutions existent ; dans le cadre de l'image, une approche multi-échelle (plusieurs entrées à différentes échelles de la même région de l'image) devrait permettre d'étendre le champs de vision effectif du réseau.

L'architecture multitâche ayant montré son efficacité sur au moins deux tâches, il devrait être intéressant de la diversifier encore plus. L'ajout de modules de segmentation des marqueurs sains (papille, veines/artères, macula) fournirait une architecture tout-en-un. La représentation interne des caractéristiques apprises par le réseau devrait dans ce cas de figure être particulièrement intéressante, car elle condenserait de façon compressée la représentation complète de la rétine.

Le type de modèle multitâche utilisé dans nos travaux appartient à une certaine catégorie : le *Hard Parameter Sharing* (partage des poids avec contrainte). Concrètement, cela signifie qu'un module encodeur est partagé par plusieurs décodeurs indépendants. L'encodeur est par conséquent entraîné par les gradients en provenance de chacun des modules de décodage. L'indépendance des décodeurs n'est pas forcément souhaitable : il serait pertinent de rajouter un conditionnement entre les différentes branches (il n'est pas possible de trouver une hémorragie au milieu du disque optique, ce que le réseau dans son stade actuel ne peut apprendre). En pratique, le *Soft Parameter Sharing*, bien que théorisé, ne fait pas l'objet de nombreux travaux ; il s'agit donc là d'un axe de recherche novateur.

La segmentation sémantique actuelle est très simpliste (lésions rouges ou blanches). L'étape suivante est naturellement de raffiner cette classification pour une identification plus précise des lésions. Cependant, à ce stade, le manque de données liées à ces tâches restreint le champs de recherche.

L'objectif final du travail de segmentation est de permettre d'arriver à un dépistage de maladie robuste et transparent. Pour cela, des règles de diagnostic en fonction des lésions détectées sont à établir. De futurs travaux se concentreront nécessairement sur cette question, ouvrant la voie à des applications de dépistage. Le système complet serait ainsi en mesure de fournir un diagnostic et de le justifier avec les lésions repérées.

CHAPITRE 7 CONCLUSION

Le travail effectué au long de cette maîtrise est un pas de plus vers des outils de dépistage automatique plus « intelligents », capables de détecter les traces de plusieurs maladies. A la différence d'autres approches existantes, la volonté de construire ce système en se basant sur des cartes de segmentations de divers marqueurs joue en faveur de la transparence dans la prise de décision.

A cette fin, l'architecture de réseaux de neurones proposée pourra servir de référence, voire même être étendue en lui ajoutant d'autres tâches dont la segmentation des marqueurs anatomiques sains.

L'intégration de ce travail de recherche dans le cadre d'un projet plus large et ambitieux va lui permettre d'être évalué dans des conditions de terrains plus réalistes que les évaluations usuelles de performance sur des bases de données relativement standards et surtout limitées en taille. Les objectifs de développement étant fixés pour la plateforme à développer, son déploiement devrait pouvoir s'opérer progressivement dans un futur proche. La création d'une base de données élargie et précisément annotée ouvrira la porte à de nombreuses perspectives de recherches qui à terme permettront de s'approcher à un niveau de performance similaire à un clinicien expert.

RÉFÉRENCES

“Ai diagnostics need attention”, Nature, vol. 555, p. 285, Mars 2018. DOI : 10.1038/d41586-018-03067-x. En ligne : <https://www.nature.com/magazine-assets/d41586-018-03067-x/d41586-018-03067-x.pdf>

M. D. Abramoff, W. L. Alward, E. C. Greenlee, L. Shuba, C. Y. Kim, J. H. Fingert, et Y. H. Kwon, “Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features”, Investigative ophthalmology & visual science, vol. 48, no. 4, pp. 1665–1673, 2007.

M. D. Abràmoff, M. K. Garvin, et M. Sonka, “Retinal imaging and image analysis”, IEEE Rev Biomed Eng, vol. 3, pp. 169–208, Jan 2010, 22275207[pmid]. DOI : 10.1109/RBME.2010.2084567. En ligne : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3131209/>

B. Al-Bander, M. B. Williams, W. Al-Nuaimy, A. M. Al-Tae, H. Pratt, et Y. Zheng, Dense Fully Convolutional Segmentation of the Optic Disc and Cup in Colour Fundus for Glaucoma Diagnosis, 2018, vol. 10, no. 4. DOI : 10.3390/sym10040087

A. Aquino, M. E. Gegundez-Arias, et D. Marin, “Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques”, IEEE Transactions on Medical Imaging, vol. 29, no. 11, pp. 1860–1869, Nov 2010. DOI : 10.1109/TMI.2010.2053042

H. Bay, A. Ess, T. Tuytelaars, et L. V. Gool, “Speeded-up robust features (surf)”, Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346 – 359, 2008, similarity Matching in Computer Vision and Multimedia. DOI : <https://doi.org/10.1016/j.cviu.2007.09.014>. En ligne : <http://www.sciencedirect.com/science/article/pii/S1077314207001555>

U. Budak, A. Şengür, Y. Guo, et Y. Akbulut, “A novel microaneurysms detection approach based on convolutional neural networks with reinforcement sample learning algorithm”, Health Information Science and Systems, vol. 5, no. 1, p. 14, Nov 2017. DOI : 10.1007/s13755-017-0034-9. En ligne : <https://doi.org/10.1007/s13755-017-0034-9>

J. S. Cardoso et L. Corte-Real, “Toward a generic evaluation of image segmentation”, IEEE Transactions on Image Processing, vol. 14, no. 11, pp. 1773–1782, 2005.

- E. J. Carmona, M. Rincón, J. García-Feijoó, et J. M. M. de-la Casa, “Identification of the optic nerve head with genetic algorithms”, Artificial Intelligence in Medicine, vol. 43, no. 3, pp. 243 – 259, 2008. DOI : <http://dx.doi.org/10.1016/j.artmed.2008.04.005>. En ligne : <http://www.sciencedirect.com/science/article/pii/S09333365708000547>
- A. Cauchy, “Méthode générale pour la résolution des systemes d’équations simultanées”, Comp. Rend. Sci. Paris, vol. 25, no. 1847, pp. 536–538, 1847.
- P. Chudzik, M. Somshubra, F. Caliva, B. Al-Diri, A. Hunter et al., “Exudate segmentation using fully convolutional neural networks and inception modules”. SPIE, 2018.
- J. Cohen, “A coefficient of agreement for nominal scales”, Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, 1960. DOI : 10.1177/001316446002000104. En ligne : <https://doi.org/10.1177/001316446002000104>
- M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, et J. V. Forrester, “A fully automated comparative microaneurysm digital detection system”, Eye, vol. 11, no. 5, p. 622, 1997.
- G. Cybenko, “Approximation by superpositions of a sigmoidal function”, Mathematics of Control, Signals and Systems, vol. 2, no. 4, pp. 303–314, Dec 1989. DOI : 10.1007/BF02551274. En ligne : <https://doi.org/10.1007/BF02551274>
- E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, B. Lay, et A. Chabouis, “Teleophta : Machine learning and image processing methods for teleophthalmology”, IRBM, vol. 34, no. 2, pp. 196 – 203, 2013, special issue : ANR TECSAN : Technologies for Health and Autonomy. DOI : <https://doi.org/10.1016/j.irbm.2013.01.010>. En ligne : <http://www.sciencedirect.com/science/article/pii/S1959031813000237>
- E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, et J.-C. Klein, “Feedback on a publicly distributed image database : The messidor database”, Image Analysis & Stereology, vol. 33, no. 3, pp. 231–234, 2014. DOI : 10.5566/ias.1155. En ligne : <https://www.ias-iss.org/ojs/IAS/article/view/1155>
- A. P. Dempster, N. M. Laird, et D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm”, Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, pp. 1–38, 1977. En ligne : <http://www.jstor.org/stable/2984875>

L. R. Dice, “Measures of the amount of ecologic association between species”, Ecology, vol. 26, no. 3, pp. 297–302, 1945.

Early Treatment Diabetic Retinopathy Study Research Group et al., “Early treatment diabetic retinopathy study design and baseline patient characteristics : Etdrs report number 7”, Ophthalmology, vol. 98, no. 5, pp. 741–756, 1991.

J. D. Fauw, “Detecting diabetic retinopathy in eye images”, Juil. 2015. En ligne : <http://jeffreydf.github.io/diabetic-retinopathy-detection/>

FDA, “Fda permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems”, FDA News Release, Avr. 2018, u.S. Food and Drug Administration. En ligne : <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm>

A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, et P. F. Sharp, “Automated assessment of diabetic retinal image quality based on clarity and field definition”, Investigative Ophthalmology & Visual Science, vol. 47, no. 3, p. 1120, 2006. DOI : 10.1167/iovs.05-1155. En ligne : [+http://dx.doi.org/10.1167/iovs.05-1155](http://dx.doi.org/10.1167/iovs.05-1155)

M. Foracchia, E. Grisan, et A. Ruggeri, “Detection of optic disc in retinal images by means of a geometrical model of vessel structure”, IEEE Transactions on Medical Imaging, vol. 23, no. 10, pp. 1189–1195, Oct 2004. DOI : 10.1109/TMI.2004.829331

A. F. Frangi, W. J. Niessen, K. L. Vincken, et M. A. Viergever, “Multiscale vessel enhancement filtering”, dans Medical Image Computing and Computer-Assisted Intervention — MICCAI’98, W. M. Wells, A. Colchester, et S. Delp, édés. Berlin, Heidelberg : Springer Berlin Heidelberg, 1998, pp. 130–137.

F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, et M. Gonzalez-Hernandez, “Rim-one : An open retinal image database for optic nerve evaluation”, dans 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), June 2011, pp. 1–6. DOI : 10.1109/CBMS.2011.5999143

I. Goodfellow, Y. Bengio, et A. Courville, Deep Learning. MIT Press, 2016, <http://www.deeplearningbook.org>.

M. Haloi, “Improved microaneurysm detection using deep neural networks”, CoRR, vol. abs/1505.04424, 2015. En ligne : <http://arxiv.org/abs/1505.04424>

Dépistage et diagnostic précoce du Glaucome : Problématique et perspectives en France, HAS (Haute Autorité de Santé), Nov. 2006.

K. He, X. Zhang, S. Ren, et J. Sun, “Deep residual learning for image recognition”, dans 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 770–778. DOI : 10.1109/CVPR.2016.90

B. Herman, W. Peter, et M. Lene, “Automated quality evaluation of digital fundus photographs”, Acta Ophthalmologica, vol. 87, no. 6, pp. 643–647, 2009. DOI : 10.1111/j.1755-3768.2008.01321.x. En ligne : <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1755-3768.2008.01321.x>

A. Hoover et M. Goldbaum, “Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels”, IEEE Transactions on Medical Imaging, vol. 22, no. 8, pp. 951–958, Aug 2003. DOI : 10.1109/TMI.2003.815900

A. D. Hoover, V. Kouznetsova, et M. Goldbaum, “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response”, IEEE Transactions on Medical Imaging, vol. 19, no. 3, pp. 203–210, March 2000. DOI : 10.1109/42.845178

D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito et al., “Optical coherence tomography”, Science, vol. 254, no. 5035, pp. 1178–1181, 1991.

Q. Huang et B. Dom, “Quantitative methods of evaluating image segmentation”, dans Image processing, 1995. proceedings., international conference on, vol. 3. IEEE, 1995, pp. 53–56.

P. Jaccard, “Etude de la distribution florale dans une portion des alpes et du jura”, vol. 37, pp. 547–579, 01 1901.

S. Joshi et P. Karule, “A review on exudates detection methods for diabetic retinopathy”, Biomedicine & Pharmacotherapy, vol. 97, pp. 1454 – 1460, 2018. DOI : <https://doi.org/10.1016/j.biopha.2017.11.009>. En ligne : <http://www.sciencedirect.com/science/article/pii/S0753332217346267>

J. Kaluzny, H. Li, W. Liu, P. Nesper, J. Park, H. F. Zhang, et A. A. Fawzi, “Bayer filter snapshot hyperspectral fundus camera for human retinal imaging”, Current Eye Research, vol. 42, no. 4, pp. 629–635, 2017, pMID : 27767345. DOI : 10.1080/02713683.2016.1221976. En ligne : <https://doi.org/10.1080/02713683.2016.1221976>

T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kalviainen, et J. Pietia, “Diaretdb1 diabetic retinopathy database and evaluation protocol”, 01 2007.

J. Kaur et D. Mittal, “A generalized method for the segmentation of exudates from pathological retinal fundus images”, Biocybernetics and Biomedical Engineering, vol. 38, no. 1, pp. 27 – 53, 2018. DOI : <https://doi.org/10.1016/j.bbe.2017.10.003>. En ligne : <http://www.sciencedirect.com/science/article/pii/S0208521617302711>

A. Khawaja, I. Usman, et A. Khan, “Automated fundoscopic recognition of diabetic retinopathy features : a review”, dans Informatics, Health & Technology (ICIHT), International Conference on. IEEE, 2017, pp. 1–7.

D. P. Kingma et J. Ba, “Adam : A method for stochastic optimization”, arXiv preprint arXiv :1412.6980, 2014.

G. Klambauer, T. Unterthiner, A. Mayr, et S. Hochreiter, “Self-normalizing neural networks”, CoRR, vol. abs/1706.02515, 2017. En ligne : <http://arxiv.org/abs/1706.02515>

S. Kullback et R. A. Leibler, “On information and sufficiency”, The annals of mathematical statistics, vol. 22, no. 1, pp. 79–86, 1951.

T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilík, G. Michelson, et J. Hornegger, “Automatic no-reference quality assessment for retinal fundus images using vessel segmentation”, dans Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, June 2013, pp. 95–100. DOI : [10.1109/CBMS.2013.6627771](https://doi.org/10.1109/CBMS.2013.6627771)

M. Lalonde, M. Beaulieu, et L. Gagnon, “Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching”, IEEE Transactions on Medical Imaging, vol. 20, no. 11, pp. 1193–1200, Nov 2001. DOI : [10.1109/42.963823](https://doi.org/10.1109/42.963823)

Y. LeCun, L. Bottou, Y. Bengio, et P. Haffner, “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, Nov 1998. DOI : [10.1109/5.726791](https://doi.org/10.1109/5.726791)

Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, et L. D. Jackel, “Handwritten digit recognition with a back-propagation network”, dans Advances in neural information processing systems, 1990, pp. 396–404.

C. Y. Lee, P. Gallagher, et Z. Tu, “Generalizing pooling functions in cnns : Mixed, gated, and tree”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 863–875, April 2018. DOI : 10.1109/TPAMI.2017.2703082

L. Likforman-Sulem et E. Barney-Smith, Reconnaissance des formes, ellipses, éd. Technosup, Oct. 2013.

Q. Liu, B. Zou, J. Chen, W. Ke, K. Yue, Z. Chen, et G. Zhao, “A location-to-segmentation strategy for automatic exudate segmentation in colour retinal fundus images”, Computerized Medical Imaging and Graphics, vol. 55, pp. 78 – 86, 2017, special Issue on Ophthalmic Medical Image Analysis. DOI : <https://doi.org/10.1016/j.compmedimag.2016.09.001>. En ligne : <http://www.sciencedirect.com/science/article/pii/S0895611116300933>

J. Long, E. Shelhamer, et T. Darrell, “Fully convolutional networks for semantic segmentation”, dans Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, et L. Kennedy, “Optic nerve head segmentation”, IEEE Transactions on Medical Imaging, vol. 23, no. 2, pp. 256–264, Feb 2004. DOI : 10.1109/TMI.2003.823261

E. B. M., H. O. K., L. O. V., et B. Toke, “The relationship between age and colour content in fundus images”, Acta Ophthalmologica Scandinavica, vol. 80, no. 5, pp. 485–489. DOI : 10.1034/j.1600-0420.2002.800505.x. En ligne : <https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0420.2002.800505.x>

W. S. McCulloch et W. Pitts, “A logical calculus of the ideas immanent in nervous activity”, The bulletin of mathematical biophysics, vol. 5, no. 4, pp. 115–133, 1943.

D. McLeod, “Reappraisal of the retinal cotton-wool spot : a discussion paper.” J R Soc Med, vol. 74, no. 9, pp. 682–686, Sep 1981, 6169833[pmid]. En ligne : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1438890/>

S. Morales, V. Naranjo, J. Angulo, et M. Alcañiz, “Automatic detection of optic disc based on pca and mathematical morphology”, IEEE Transactions on Medical Imaging, vol. 32, no. 4, pp. 786–796, April 2013. DOI : 10.1109/TMI.2013.2238244

U. T. Nguyen, A. Bhuiyan, L. A. Park, et K. Ramamohanarao, “An effective retinal blood vessel segmentation method using multi-scale line detection”, Pattern Recognition, vol. 46,

no. 3, pp. 703 – 715, 2013. DOI : <https://doi.org/10.1016/j.patcog.2012.08.009>.
En ligne : <http://www.sciencedirect.com/science/article/pii/S003132031200355X>

M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, et M. D. Abramoff, “Automatic detection of red lesions in digital color fundus photographs”, IEEE Transactions on Medical Imaging, vol. 24, no. 5, pp. 584–592, May 2005. DOI : 10.1109/TMI.2005.843738

M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quéllec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, et M. D. Abramoff, “Retinopathy online challenge : Automatic detection of microaneurysms in digital color fundus photographs”, IEEE Transactions on Medical Imaging, vol. 29, no. 1, pp. 185–195, Jan 2010. DOI : 10.1109/TMI.2009.2033909

M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. van Ginneken, J. C. Folk, et M. D. Abramoff, “Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs”, IEEE Transactions on Medical Imaging, vol. 30, no. 11, pp. 1941–1950, Nov 2011. DOI : 10.1109/TMI.2011.2159619

M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. A. Suttorp-Schulten, et M. D. Abramoff, “Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis”, Investigative Ophthalmology & Visual Science, vol. 48, no. 5, p. 2260, 2007. DOI : 10.1167/iovs.06-0996. En ligne : [+http://dx.doi.org/10.1167/iovs.06-0996](http://dx.doi.org/10.1167/iovs.06-0996)

I. Olkin et F. Pukelsheim, “The distance between two random vectors with given dispersion matrices”, Linear Algebra and its Applications, vol. 48, pp. 257 – 263, 1982. DOI : [https://doi.org/10.1016/0024-3795\(82\)90112-4](https://doi.org/10.1016/0024-3795(82)90112-4). En ligne : <http://www.sciencedirect.com/science/article/pii/0024379582901124>

S. Onal, X. Chen, V. Satamraju, M. Balasooriya, et H. Dabil-Karacal, “Automated and simultaneous fovea center localization and macula segmentation using the new dynamic identification and classification of edges model”, J Med Imaging (Bellingham), vol. 3, no. 3, p. 034002, Jul 2016, 16112R[PII]. DOI : 10.1117/1.JMI.3.3.034002. En ligne : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5019108/>

C. Peng, X. Zhang, G. Yu, G. Luo, et J. Sun, “Large kernel matters - improve semantic segmentation by global convolutional network”, CoRR, vol. abs/1703.02719, 2017. En ligne : <http://arxiv.org/abs/1703.02719>

F. Perronnin, J. Sánchez, et T. Mensink, “Improving the fisher kernel for large-scale image classification”, dans Computer Vision – ECCV 2010, K. Daniilidis, P. Maragos, et N. Paragios, édés. Berlin, Heidelberg : Springer Berlin Heidelberg, 2010, pp. 143–156.

S. Philip, L. M. Cowie, et J. A. Olson, “The impact of the health technology board for scotland’s grading model on referrals to ophthalmology services”, Br J Ophthalmol, vol. 89, no. 7, pp. 891–896, Jul 2005, 0890891[PII]. DOI : 10.1136/bjo.2004.051334. En ligne : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1772738/>

R. Pires, S. Avila, H. F. Jelinek, J. Wainer, E. Valle, et A. Rocha, “Beyond lesion-based diabetic retinopathy : A direct approach for referral”, IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 193–200, Jan 2017. DOI : 10.1109/JBHI.2015.2498104

C. Playout, R. Duval, et F. Cheriet, “A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images”, dans Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018, 2018.

B. T. Polyak, “Some methods of speeding up the convergence of iteration methods”, USSR Computational Mathematics and Mathematical Physics, vol. 4, no. 5, pp. 1–17, 1964.

G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, et C. Roux, “Optimal wavelet transform for the detection of microaneurysms in retina photographs”, IEEE Transactions on Medical Imaging, vol. 27, no. 9, pp. 1230–1241, Sept 2008. DOI : 10.1109/TMI.2008.920619

G. Quellec, K. Charrière, Y. Boudi, B. Cochener, et M. Lamard, “Deep image mining for diabetic retinopathy screening”, Medical Image Analysis, vol. 39, pp. 178 – 193, 2017. DOI : <https://doi.org/10.1016/j.media.2017.04.012>. En ligne : <http://www.sciencedirect.com/science/article/pii/S136184151730066X>

J. Rahebi et F. Hardalaç, “A new approach to optic disc detection in human retinal images using the firefly algorithm”, Medical & Biological Engineering & Computing, vol. 54, no. 2, pp. 453–461, Mar 2016. DOI : 10.1007/s11517-015-1330-7. En ligne : <https://doi.org/10.1007/s11517-015-1330-7>

O. Ronneberger, P. Fischer, et T. Brox, “U-net : Convolutional networks for biomedical image segmentation”, dans Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, et A. F. Frangi, édés. Cham : Springer International Publishing, 2015, pp. 234–241.

P. Roy, R. Tennakoon, K. Cao, S. Sedai, D. Mahapatra, S. Maetschke, et R. Garnavi, “A novel hybrid approach for severity assessment of diabetic retinopathy in colour fundus images”, dans 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), April 2017, pp. 1078–1082. DOI : 10.1109/ISBI.2017.7950703

S. Ruder, “An overview of multi-task learning in deep neural networks”, CoRR, vol. abs/1706.05098, 2017. En ligne : <http://arxiv.org/abs/1706.05098>

M. Schmitt, “Content-based image enhancement for eye fundus images visualization”, Mémoire de maîtrise, Université de Montréal, 2017.

L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet, et J. M. P. Langlois, “Red lesion detection using dynamic shape features for diabetic retinopathy screening”, IEEE Transactions on Medical Imaging, vol. 35, no. 4, pp. 1116–1126, April 2016. DOI : 10.1109/TMI.2015.2509785

C. Sinthanayothin, J. Boyce, H. Cook, et T. Williamson, “Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images”, Br J Ophthalmol, vol. 83, no. 8, pp. 902–910, Aug 1999, 10413690[pmid]. En ligne : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1723142/>

C. Sisson, S. Farnand, M. Fairchild, et B. Fischer, “Analysis of color consistency in retinal fundus photography : Application of color management and development of an eye model standard”, vol. 2014, pp. 1–2, 12 2014.

J. Sivaswamy, S. Krishnadas, A. Chakravarty, G. Joshi, A. S. Tabish et al., “A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis”, JSM Biomedical Imaging Data Papers, vol. 2, no. 1, p. 1004, 2015.

K. A. Spackman, “Signal detection theory : Valuable tools for evaluating inductive learning”, dans Proceedings of the sixth international workshop on Machine learning. Elsevier, 1989, pp. 160–163.

T. Spencer, R. P. Phillips, P. F. Sharp, et J. V. Forrester, “Automated detection and quantification of microaneurysms in fluorescein angiograms”, Graefes Archive for Clinical and Experimental Ophthalmology, vol. 230, no. 1, pp. 36–41, Jan 1992. DOI : 10.1007/BF00166760. En ligne : <https://doi.org/10.1007/BF00166760>

S. Sreng, J. i. Takada, N. Maneerat, D. Isarakorn, B. Pasaya, R. Panjaphongse, et R. Varakulsiripunth, “Automatic exudate extraction for early detection of diabetic retinopathy”,

dans 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), Oct 2013, pp. 31–35. DOI : 10.1109/ICITEED.2013.6676206

J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, et B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina”, IEEE Transactions on Medical Imaging, vol. 23, no. 4, pp. 501–509, April 2004. DOI : 10.1109/TMI.2004.825627

I. Sutskever, J. Martens, G. Dahl, et G. Hinton, “On the importance of initialization and momentum in deep learning”, dans International conference on machine learning, 2013, pp. 1139–1147.

L. Tang, M. K. Garvin, K. Lee, W. L. W. Alward, Y. H. Kwon, et M. D. Abramoff, “Robust multiscale stereo matching from fundus images with radiometric differences”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 11, pp. 2245–2258, Nov 2011. DOI : 10.1109/TPAMI.2011.69

A. M. Taori, A. K. Chaudhari, S. S. Patankar, et J. V. Kulkarni, “Segmentation of macula in retinal images using automated seeding region growing technique”, dans 2016 International Conference on Inventive Computation Technologies (ICICT), vol. 2, Aug 2016, pp. 1–5. DOI : 10.1109/INVENTIVE.2016.7824792

D. Usher, M. Himaga, M. Dumskyj, et J. Boyce, “Automated assessment of digital fundus image quality using detected vessel area”, dans Proceedings of Medical Image Understanding and Analysis. Citeseer, 2003, pp. 81–84.

G. V., P. L., C. M., et al, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs”, JAMA, vol. 316, no. 22, pp. 2402–2410, 2016. DOI : 10.1001/jama.2016.17216. En ligne : [+http://dx.doi.org/10.1001/jama.2016.17216](http://dx.doi.org/10.1001/jama.2016.17216)

M. J. van Grinsven, T. Theelen, L. Witkamp, J. van der Heijden, J. P. van de Ven, C. B. Hoyng, B. van Ginneken, et C. I. Sánchez, “Automatic differentiation of color fundus images containing drusen or exudates using a contextual spatial pyramid approach”, Biomedical optics express, vol. 7, no. 3, pp. 709–725, 2016.

C. Van Rijsbergen et D. C. Blair, “Information retrieval”, Journal of the American Society for Information Science, vol. 30, no. 6, pp. 374–375, 1979.

T. Walter, J. C. Klein, P. Massin, et A. Erginay, “A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human

retina”, IEEE Transactions on Medical Imaging, vol. 21, no. 10, pp. 1236–1243, Oct 2002. DOI : 10.1109/TMI.2002.806290

T. Walter et J.-C. Klein, “Segmentation of color fundus images of the human retina : Detection of the optic disc and the vascular tree using morphological techniques”, dans Medical Data Analysis, J. Crespo, V. Maojo, et F. Martin, éd. Berlin, Heidelberg : Springer Berlin Heidelberg, 2001, pp. 282–287.

S. Wang, K. Jin, H. Lu, C. Cheng, J. Ye, et D. Qian, “Human visual system-based fundus image quality assessment of portable fundus camera photographs”, IEEE Transactions on Medical Imaging, vol. 35, no. 4, pp. 1046–1055, April 2016. DOI : 10.1109/TMI.2015.2506902

Wikipedia, “Hypertensive retinopathy — Wikipedia, the free encyclopedia”, 2018, [Online ; accessed 24-July-2018]. En ligne : https://en.wikipedia.org/w/index.php?title=Hypertensive_retinopathy&oldid=837926454

Wikipédia, “Œil humain — wikipédia, l’encyclopédie libre”, 2018, [En ligne ; Page disponible le 22-avril-2018]. En ligne : http://fr.wikipedia.org/w/index.php?title=%C5%92il_humain&oldid=147767737

D. W. K. Wong, J. Liu, J. H. Lim, X. Jia, F. Yin, H. Li, et T. Y. Wong, “Level-set based automatic cup-to-disc ratio determination using retinal fundus images in argali”, dans 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug 2008, pp. 2266–2269. DOI : 10.1109/IEMBS.2008.4649648

J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund et al., “Global prevalence and major risk factors of diabetic retinopathy”, Diabetes care, vol. 35, no. 3, pp. 556–564, 2012.

A. A. H. A. R. Youssif, A. Z. Ghalwash, et A. A. S. A. R. Ghoneim, “Optic disc detection from normalized digital fundus images by means of a vessels’ direction matched filter”, IEEE Transactions on Medical Imaging, vol. 27, no. 1, pp. 11–18, Jan 2008. DOI : 10.1109/TMI.2007.900326

X. Zhang, G. Thibault, E. Decenci re, B. Marcotegui, B. La y, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin, A. Chabouis, Z. Victor, et A. Erginay, “Exudate detection in color retinal images for mass screening of diabetic retinopathy”, Medical Image Analysis, vol. 18, no. 7, pp. 1026 –

1043, 2014. DOI : <https://doi.org/10.1016/j.media.2014.05.004>. En ligne : <http://www.sciencedirect.com/science/article/pii/S1361841514000693>

Y. Zhang, “A survey on evaluation methods for image segmentation”, Pattern Recognition, vol. 29, no. 8, pp. 1335 – 1346, 1996. DOI : [https://doi.org/10.1016/0031-3203\(95\)00169-7](https://doi.org/10.1016/0031-3203(95)00169-7). En ligne : <http://www.sciencedirect.com/science/article/pii/0031320395001697>

Y. Zheng, E. L. Lamoureux, R. Lavanya, R. Wu, M. K. Ikram, J. J. Wang, P. Mitchell, N. Cheung, T. Aung, S.-M. Saw, et T. Y. Wong, “Prevalence and risk factors of diabetic retinopathy in migrant indians in an urbanized society in asia : The singapore indian eye study”, Ophthalmology, vol. 119, no. 10, pp. 2119 – 2124, 2012. DOI : <https://doi.org/10.1016/j.ophtha.2012.04.027>. En ligne : <http://www.sciencedirect.com/science/article/pii/S0161642012003995>